



# On absorbtion times and Dirichlet eigenvalues

Laurent Miclo

## ► To cite this version:

Laurent Miclo. On absorbtion times and Dirichlet eigenvalues. ESAIM: Probability and Statistics, 2010, 14, pp.117-150. 10.1051/ps:2008037 . hal-00270958

**HAL Id: hal-00270958**

**<https://hal.science/hal-00270958>**

Submitted on 8 Apr 2008

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# On absorption times and Dirichlet eigenvalues

Laurent Miclo

Laboratoire d'Analyse, Topologie, Probabilités, U.M.R. 6632  
Université de Provence and C.N.R.S.  
France

## Abstract

This paper gives a stochastic representation in spectral terms for the absorption time  $T$  of a finite Markov chain which is irreducible and reversible outside the absorbing point. This yields quantitative informations on the parameters of a similar representation due to O'Cinneide for general chains admitting real eigenvalues. In the discrete time setting, if the underlying Dirichlet eigenvalues (namely the eigenvalues of the Markov transition operator restricted to the functions vanishing on the absorbing point) are nonnegative, we show that  $T$  is distributed as a mixture of sums of independent geometric laws whose parameters are successive Dirichlet eigenvalues (starting from the smallest one). The mixture weights depend on the starting law. This result leads to a probabilistic interpretation of the spectrum, in terms of strong random times and local equilibria through a simple intertwining relation. Next this study is extended to the continuous time framework, where geometric laws have to be replaced by exponential distributions having the (opposite) Dirichlet eigenvalues of the generator as parameters. Returning to the discrete time setting we consider the influence of negative eigenvalues which are given another probabilistic meaning. These results generalize results of Karlin and McGregor and Keilson for birth and death chains.

**Keywords:** irreducible and reversible subMarkovian matrices, exit or absorption times, Dirichlet eigenvalues, mixtures, geometric laws, exponential distributions, strong random times, local equilibria, intertwining, birth and death chains and processes.

**MSC2000:** first: 60J10 secondary: 60J27, 37A30, 31C25, 60J80.

# 1 Introduction

This paper shows that the time to absorption of a reversible Markov chain has a stochastic representation in spectral terms. It gives quantitative informations on the parameters of a similar representation due to O’Cinneide [24] for general chains admitting real eigenvalues. In continuous time, the absorption time is distributed as a mixture of sums of independent exponentials having the underlying eigenvalues as parameters. This generalizes results of Karlin and McGregor [15] and Keilson [17] for birth and death chains.

We begin with the case of discrete time. Let  $\bar{S}$  be a finite set endowed with a probability transition matrix  $\bar{P}$ . We make the following irreducibility and reversibility assumptions:

**(A1):** There exists a particular point  $\triangle \in \bar{S}$  which is absorbing (i.e.  $\bar{P}(\triangle, \triangle) = 1$ ) and such that  $\{\triangle\}$  is the unique irreducible class of  $\bar{P}$  (or equivalently, for any  $x \in \bar{S}$ , there exists a path going from  $x$  to  $\triangle$ ,  $x = x_0, x_1, \dots, x_p = \triangle$ , satisfying  $\bar{P}(x_i, x_{i+1}) > 0$  for  $i \in \llbracket 0, p-1 \rrbracket$ ). Let  $S := \bar{S} \setminus \{\triangle\}$ , we assume furthermore that the subMarkovian restriction  $P := (\bar{P}(x, y))_{x, y \in S}$  of  $\bar{P}$  to  $S$  is irreducible (namely, any pair of points from  $S$  can be joined by a path).

**(A2):**  $P$  is reversible with respect to some positive probability  $\pi$  on  $S$ :

$$\forall x, y \in S, \quad \pi(x)P(x, y) = \pi(y)P(y, x)$$

Let  $m_0$  be a probability on  $S$ , we consider  $X := (X_n)_{n \in \mathbb{N}}$  a homogeneous Markov chain on  $\bar{S}$  whose initial distribution is  $m_0$  and whose transitions are given by  $\bar{P}$ . We are interested in the absorption time

$$T := \inf\{n \in \mathbb{N} : X_n = \triangle\}$$

which is almost surely finite by the first irreducibility hypothesis.

To describe its law, we need to introduce the corresponding Dirichlet eigenvalues, which are the eigenvalues of the restriction of  $\bar{P}$  to the space of functions  $f : \bar{S} \rightarrow \mathbb{R}$  satisfying the Dirichlet condition  $f(\triangle) = 0$ . Equivalently, they are the eigenvalues of  $P$ ,

$$1 > \theta_0 > \theta_1 \geq \theta_2 \geq \dots \geq \theta_{N-1} > -1$$

where  $N = \text{card}(S)$ . The fact that they are real numbers comes from the reversibility assumption. They belong to  $(-1, 1)$  because  $\{\triangle\}$  is the unique irreducible class of  $\bar{P}$ . The inequality  $\theta_0 > \theta_1$  (and also  $\theta_0 \geq |\theta_{N-1}|$ ) is a consequence of the irreducibility of  $P$  and of the Perron-Frobenius theorem.

To get a simple statement below, we assume that all these eigenvalues are nonnegative:

**(A3):**  $\theta_{N-1} \geq 0$ .

For  $0 \leq i \leq N-1$ , we denote by  $\mathcal{G}(\theta_i, \theta_{i+1}, \dots, \theta_{N-1})$  the convolution of geometric distributions of parameters  $\theta_i, \theta_{i+1}, \dots, \theta_{N-1}$ . More precisely, this is the law of  $\tau_i + \tau_{i+1} + \dots + \tau_{N-1}$ , where the  $\tau_j$ , for  $j \in \llbracket i, N-1 \rrbracket$ , are independent random variables and satisfy

$$\forall j \in \llbracket i, N-1 \rrbracket, \forall n \in \mathbb{N}^*, \quad \mathbb{P}[\tau_j = n] = \theta_j^{n-1}(1 - \theta_j)$$

If we denote  $\theta := (\theta_0, \dots, \theta_{N-1})$  and if  $a := (a_0, \dots, a_{N-1})$  is a probability on  $\llbracket 0, N-1 \rrbracket$ , we introduce  $\mathcal{M}_{\mathcal{G}}(\theta, a)$  the mixture of the distributions  $\mathcal{G}(\theta_i, \theta_{i+1}, \dots, \theta_{N-1})$ , for  $i \in \llbracket 0, N-1 \rrbracket$ , with respective weights given by  $a$ , i.e.

$$\mathcal{M}_{\mathcal{G}}(\theta, a) := \sum_{0 \leq i \leq N-1} a_i \mathcal{G}(\theta_i, \theta_{i+1}, \dots, \theta_{N-1})$$

From a probabilistic point of view, if  $T_i$ , for  $i \in \llbracket 0, N-1 \rrbracket$ , are respectively distributed according to  $\mathcal{G}(\theta_i, \theta_{i+1}, \dots, \theta_{N-1})$  and if  $A$  is independent of  $(T_i)_{i \in \llbracket 0, N-1 \rrbracket}$  and has  $a$  as law, then  $T_A$  has distribution  $\mathcal{M}_{\mathcal{G}}(\theta, a)$ .

One of our main results can now be stated (see the paragraph after Theorem 2 below for a discussion of the links with the representation due to O’Cinneide [24]):

**Theorem 1** *Under the assumptions (A1), (A2) and (A3), there exists a probability  $a := (a_0, \dots, a_{N-1})$ , with  $a_0 > 0$ , such that  $T$  is distributed as  $\mathcal{M}_G(\theta, a)$ .*

The probability  $a$  strongly depends on the initial distribution  $m_0$  of  $X$ . Let us write  $a(m_0)$ , then  $a$  commutes with the operation of taking mixtures: if  $m'_0$  and  $m''_0$  are two probabilities on  $S$  and  $\alpha \in [0, 1]$ , then  $a(\alpha m'_0 + (1 - \alpha)m''_0) = \alpha a(m'_0) + (1 - \alpha)a(m''_0)$ . Thus to prove the above theorem, it is sufficient to deal with the case where  $m_0$  is a Dirac mass at some initial point  $x \in S$ , nevertheless it will be more fruitful to treat the general situation directly. In the same spirit, if  $P$  is not assumed to be irreducible (then the inequality  $\theta_0 > \theta_1$  is not necessarily satisfied), we can decompose  $S$  into the corresponding irreducibility classes for  $P$ , say  $S_1, \dots, S_r$ , and apply the previous theorem to the Markovian transition matrices  $\bar{P}_i := (\bar{P}(x, y))_{x, y \in S_i \sqcup \{\Delta\}}$ , for  $i \in \llbracket 1, r \rrbracket$ , to get a similar result (except for the assertion  $a_0 > 0$ ), with a mixture of sum of geometric variables of parameters certain successive Dirichlet eigenvalues of the  $P_i := (\bar{P}(x, y))_{x, y \in S_i}$ , for  $i \in \llbracket 1, r \rrbracket$ .

It is also possible to describe the law of  $T$  when some of the eigenvalues are negative, but the corresponding statement is more involved and we refer the reader to Section 6 below. The case of nonnegative eigenvalues is already sufficient to deduce a similar result in the continuous time setting. Let  $\bar{S} = S \sqcup \{\Delta\}$  be a finite set, endowed with  $\bar{L}$  a Markov generator matrix (i.e. the off-diagonal entries are nonnegative and the rows sum up to zero). In analogy to what we have done before, we assume that

(B1):  $\{\Delta\}$  is the unique recurrence class of  $\bar{L}$ , so in particular  $\Delta$  is an absorbing point (namely the row  $\bar{L}(\Delta, \cdot)$  is null) and the  $S \times S$  submatrix  $L$  of  $\bar{L}$  (or from a functional point of view, the restriction of the operator  $\bar{L}$  to functions vanishing in  $\Delta$ ) is irreducible.

(B2):  $L$  is reversible with respect to some positive probability  $\pi$  on  $S$ :

$$\forall x, y \in S, \quad \pi(x)L(x, y) = \pi(y)L(y, x)$$

Let  $m_0$  be a probability on  $S$ , we consider  $X := (X_t)_{t \geq 0}$  a homogeneous Markov process on  $\bar{S}$  whose initial distribution is  $m_0$ , whose infinitesimal generator is  $\bar{L}$  and whose trajectories are cadlag, namely they are right-continuous and admit left-limits. We are interested in the absorption time

$$T := \inf\{t \geq 0 : X_t = \Delta\}$$

which is almost surely finite by the irreducibility hypothesis.

To describe its law, we need to introduce the corresponding Dirichlet eigenvalues, which are the eigenvalues of the restriction of  $-\bar{L}$  to the space of functions vanishing at  $\Delta$ . They are just the eigenvalues of  $-L$ ,

$$0 < \lambda_0 < \lambda_1 \leq \dots \leq \lambda_{N-1}$$

where  $N = \text{card}(S)$ . As above, the fact that they are real numbers comes from the reversibility assumption and they are positive because  $\{\Delta\}$  is the unique recurrence class of  $\bar{L}$ . The inequality  $\lambda_0 < \lambda_1$  is a consequence of the irreducibility of  $L$  and of the Perron-Frobenius theorem. Alternatively, one can come back to the previous situation by considering the semigroup  $(\bar{P}_t)_{t \geq 0} := (\exp(t\bar{L}))_{t \geq 0}$ . Then for any  $t > 0$ ,  $\bar{P}_t$  satisfies the above discrete time hypotheses and we have for its eigenvalues

$$\forall 0 \leq i \leq N - 1, \quad \theta_i(t) = \exp(-t\lambda_i)$$

which in particular are positive:  $\theta_{N-1}(t) > 0$ .

For  $0 \leq i \leq N - 1$ , we denote by  $\mathcal{E}(\lambda_i, \lambda_{i+1}, \dots, \lambda_{N-1})$  the convolution of exponential distributions

of parameters  $\lambda_i, \lambda_{i+1}, \dots, \lambda_{N-1}$ . More precisely, this is the law of  $\tilde{\tau}_i + \tilde{\tau}_{i+1} + \dots + \tilde{\tau}_{N-1}$ , where the  $\tilde{\tau}_j$ , for  $j \in \llbracket i, N-1 \rrbracket$ , are independent r.v. and satisfy

$$\forall j \in \llbracket i, N-1 \rrbracket, \forall t \geq 0, \quad \mathbb{P}[\tilde{\tau}_j \geq t] = \exp(-\lambda_j t)$$

Again, we denote  $\lambda := (\lambda_0, \dots, \lambda_{N-1})$  and for a probability  $a := (a_0, \dots, a_{N-1})$  on  $\llbracket 0, N-1 \rrbracket$ , we introduce  $\mathcal{M}_{\mathcal{E}}(\lambda, a)$  the mixture of the distributions  $\mathcal{E}(\lambda_i, \lambda_{i+1}, \dots, \lambda_{N-1})$ , for  $i \in \llbracket 0, N-1 \rrbracket$ , with respective weights given by  $a$ , i.e.

$$\mathcal{M}_{\mathcal{E}}(\lambda, a) := \sum_{0 \leq i \leq N-1} a_i \mathcal{E}(\lambda_i, \lambda_{i+1}, \dots, \lambda_{N-1})$$

As a consequence of previous result, we get

**Theorem 2** *Under the continuous time assumptions (B1) and (B2), there exists a probability  $a := (a_0, \dots, a_{N-1})$ , with  $a_0 > 0$ , such that  $T$  is distributed as  $\mathcal{M}_{\mathcal{E}}(\theta, a)$ .*

That  $T$  is distributed as a mixture of convolutions of exponential laws is known and due to O’Cinneide [24]. Under the relaxed assumption that the eigenvalues of  $L$  are real, he showed that there exist  $M \geq N$ , a probability  $a = (a_i)_{i \in \llbracket 0, M-1 \rrbracket}$  and positive parameters  $l = (l_i)_{i \in \llbracket 0, M-1 \rrbracket}$  such the law of  $T$  is described by

$$\mathcal{M}_{\mathcal{E}}(l, a) := \sum_{0 \leq i \leq M-1} a_i \mathcal{E}(l_i, l_{i+1}, \dots, l_{M-1})$$

O’Cinneide [24] also indicated that, once the exponential distributions have been replaced by geometric distributions, the same representation can also be deduced in discrete time by resorting to techniques from [23].

These results belong to the realm of phase type distributions (they correspond to general laws of absorbing time of finite Markov chains), which are important in the computational probability approach to queueing theory, because they lead to matrix algorithms which are tractable in practice. This field of investigation was initiated by Neuts [22] but it is out of the scope of the present article to present it. We refer the interested reader to the above references and to the bibliography therein (see also for instance Commault and Mocanu [4] or He and Zhang [13, 14] for more recent works on the subject). We just mention briefly a few features. The view point of phase type distributions is the opposite of our approach: for queueing theorists the law of  $T$  is given (this is the distribution of the arrival or service times) and they want to find a corresponding generator  $\bar{L}$  with the smallest possible state space. The information (under the reversibility assumption) that we can take  $M = N$  is important since it reduces the size of a representation in terms of a simple Markov chain. Note that in the general case of real eigenvalues, one may have to take  $M > N$ , as it is shown by an example due to Botta, Harris and Marchal [3]. Further, to estimate the minimal  $M$  seems a difficult task. The identification of the parameters as eigenvalues given in Theorem 2 is also relevant. But above all, our proof is completely different from the invariant polytope method of O’Cinneide [24] and it sheds a new probabilistic insight on his result. We hope we will be able to extend our approach (where reversibility is only needed in next section) to recover the full validity of O’Cinneide’s representation and maybe to better understand the general case where complex eigenvalues enter the game (see the nice review of O’Cinneide [25] for very motivating open problems in this direction).

An immediate application of Theorems 1 and 2, is to provide easy bounds on absorption times. Thus in continuous time (respectively in discrete time, with nonnegative eigenvalues),  $T$  is always stochastically bounded above by a sum of exponential laws (resp. geometric laws) of parameters the corresponding Dirichlet eigenvalues.

For instance, in discrete time with nonnegative eigenvalues, we get with the notation introduced before Theorem 1, for any  $n \in \mathbb{N}^*$ ,

$$\begin{aligned} \mathbb{P}[T \geq n] &\leq \mathbb{P}[\tau_1 + \dots + \tau_{N-1} \geq n] \\ &= \sum_{m \in \llbracket n, \infty \rrbracket} \sum_{n_0 + \dots + n_{N-1} = m, n_j \in \mathbb{N}^*, j \in \llbracket 0, N-1 \rrbracket} \prod (1 - \theta_j) \theta_j^{n_j - 1} \end{aligned}$$

This formula can be deduced using a divided differences computation (see also (22) in Section 6). For simplicity, assume that all the eigenvalues are distinct, then we have

$$\sum_{n_0 + \dots + n_{N-1} = m, n_j \in \mathbb{N}^*, j \in \llbracket 0, N-1 \rrbracket} \prod \theta_j^{n_j - 1} = \sum_{i \in \llbracket 0, N-1 \rrbracket} \frac{\theta_i^{N+m}}{\prod_{j \in \llbracket 0, k \rrbracket \setminus \{i\}} \theta_i - \theta_j}$$

and we recover the standard bound

$$\mathbb{P}[T \geq n] \leq \sum_{i \in \llbracket 0, N-1 \rrbracket} \frac{\theta_i^{N+n}}{\prod_{j \in \llbracket 0, k \rrbracket \setminus \{i\}} \theta_i - \theta_j}$$

Similar computations hold for continuous time. Unfortunately, these bounds are not very sharp in general, unless if  $a_0$  is not negligible. Indeed, lower bound on  $a_0$  will be important in deducing interesting quantitative informations on absorption time for specific models, but we will not investigate this subject here. Another strong motivation to study the quantity  $a_0$  is that it can be used to deduce cut-off phenomenon in the separation distance (at least in the absorbing setting instead of convergence to equilibrium, but they are related, see also a remark below). Indeed, if  $a_0$  is sufficiently close to 1 (or at least if the probability  $a$  is concentrated near 0), the strategy of Diaconis and Saloff-Coste [7] can be applied: they only work with birth and death chains because they needed the structure of the law of (dual) absorbing times. The later can also be useful for cut-off in total variation, see the recent preprint of Ding, Lubetzky and Peres [8].

Below we obtain more structural results, we present an intertwining relation between the absorbing chain or process and a special kind of chain or process, which is nonreversible. Nevertheless, the corresponding transition probability, or generator, belongs to  $\hat{\mathcal{K}}$  or  $\hat{\mathcal{L}}$ , respectively a certain kind of closure of reversible and irreducible (as before) transition probabilities with nonnegative eigenvalues, or generators, absorbing in  $\triangle$  (see sections 4 and 5 for the exact definitions). This will enable us to deduce that the class of distributions  $\mathcal{M}_{\mathcal{G}}(a, \theta)$  with  $1 > \theta_0 \geq \theta_1 \geq \dots \geq \theta_{N-1} \geq 0$  (respectively  $\mathcal{M}_{\mathcal{E}}(a, \lambda)$  with  $0 < \lambda_0 \leq \lambda_1 \leq \dots \leq \lambda_{N-1}$ ) and any probability  $a$  on  $\llbracket 0, N-1 \rrbracket$ , is exactly the set of laws of absorption times of Markov chains (resp. processes) starting from  $S$  and associated to a transition matrix belonging to  $\hat{\mathcal{K}}$  (resp. to a generator belonging to  $\hat{\mathcal{L}}$ ). This characterization goes further than the result obtained by Kent [20] in this direction, even when restricted to birth and death processes.

Theorems 1 and 2 and their proofs permit a better probabilistic understanding of the fact that under certain circumstances, the absorption time is distributed as a sum of exponential laws (or geometrical laws in discrete time with nonnegative eigenvalues) of parameters the corresponding Dirichlet eigenvalues. This is true for birth and death processes (or monotone chains) starting from one end of a finite segment and absorbing at the other end and we recover the results Karlin and McGregor [15] and Keilson [17], see for instance [6] or Fill [11] for references and the historical background on the subject.

The proof will be based on two ideas, one is the construction of strong random times (generalizing strong stationary times) in the spirit of the work of Aldous and Diaconis [1]. The other one is a stability result on symmetric matrices with nonnegative entries due to Micchelli and Willoughby [21], that we learned from two recent preprints of Fill [11, 12]. The ingredient linking these two tools is the notion of local equilibrium, which appeared in [6], but was not fully exploited there.

One limitation of our approach is that we are only able to deal with reversible chains or processes, while Theorem 1 and 2 sometimes also hold for nonreversible chains or processes, e.g. for skip-free ones, see the paper of Fill [12]. We hope to better understand this situation in future work. An easy potential extension concerns reversible Markov chains or processes with more general state space, specially diffusions (see for instance the papers [18, 19] of Kent), but this subject will not be developed here. Another even more interesting variation relates to convergence to equilibrium for reversible Markov chains or processes. Indeed, a similar approach can be applied for truly irreducible chains or processes (where the whole state space is the unique recurrence class), so that Theorems 1 and 2 equally hold for strong stationary times (for a background on this field, see for instance the paper of Aldous and Diaconis [1], Diaconis and Fill [5] or Fill [10]) with the standard (Neumann) eigenvalues. This will be the matter of a forthcoming paper.

The plan of the paper is the following: in the next section we study a particular kind of strong random time. In section 3, we take advantage of these and of a linear algebra result to define local equilibrium distributions associated to the initial probability  $m_0$ , which lead to the proof of Theorem 1. In section 4 we go further and build an intertwining relation. We deduce the similar results for continuous time in section 5 and we deal with negative eigenvalues in section 6. In the last section, we consider two antagonistic examples, one is birth and death processes and the other one is chains or processes with a constant probability or rate to go to the absorbing point.

## 2 Strong random times

In the next three sections, we will be working in the discrete time framework. This section is devoted to the construction of a special kind of strong random time which will serve as an elementary brick in building intertwining.

Let  $X := (X_n)_{n \in \mathbb{N}}$  be any homogeneous Markov chain taking values in a finite state space  $S$ . We denote by  $P$  its transition matrix and by  $m_0$  its initial distribution (seen as a row vector). By definition, a strong random time  $\tau$  is an almost surely finite randomized stopping time for  $X$  such that  $\tau$  and  $X_\tau$  are independent. Our principal result here is

**Proposition 3** *Assume there exist  $\theta \in (0, 1)$  and a probability  $\mu$  on  $S$  such that*

$$m_0 P = \theta m_0 + (1 - \theta) \mu$$

*Then there exists a strong random time  $\tau$ , distributed as a geometric law of parameter  $\theta$ , such that  $X_\tau$  admits  $\mu$  as law.*

Formally, this result also holds for  $\theta = 0$ , since we can then take  $\tau \equiv 1$ . For  $\theta = 1$  (in which case  $m_0$  is invariant for  $P$ ), the natural extension of the above statement would lead to  $\tau \equiv +\infty$ , so  $X_\tau$  cannot be defined in general, which also reflects the fact that  $\mu$  is not determined by the above equation in this situation. But if  $\mu = m_0$ , one can take  $\tau \equiv 0$ .

### Proof

The starting idea of the following construction comes from a paper of Aldous and Diaconis [1], see the proof of their Proposition 3.2.b. They were looking for a fastest strong stationary time (namely a strong random time  $\tau$  such that  $X_\tau$  is distributed according to an invariant measure for  $P$  and which is stochastically minimal among all such times).

Let us define

$$\begin{aligned} \forall x \in S, \quad s(x) &:= 1 - \frac{m_1}{\mu}(x) \\ &= \theta \left( 1 - \frac{m_0}{\mu}(x) \right) \in \mathbb{R} \sqcup \{-\infty\} \end{aligned}$$

where  $m_1 := m_0 P$  is the law of  $X_1$ . This quantity is related to the “distance” in separation between  $m_1$  and  $\mu$ , but this feature will not be explicitly used in what follows. At time 1, knowing  $X_0$  and  $X_1$ , we take  $\tau = 1$  with probability  $(1 - \theta)/(1 - s(X_1))$ , namely we generate a Bernoulli random variable with parameter  $(1 - \theta)/(1 - s(X_1))$  (conditionally independent of  $X$ , knowing the latter parameter, this is where additional noise is required) to decide if  $\tau = 1$  or not. Since  $s(X_1) \leq \theta$ , one would have checked that the previous parameter belongs to  $[0, 1]$ . So we can write

$$\mathbb{P}[\tau = 1 | X_0, X_1] = \frac{1 - \theta}{1 - s(X_1)}$$

and in particular  $\mathbb{P}[\tau = 1 | X_1] = \frac{1 - \theta}{1 - s(X_1)}$ .

Let us check that

$$\forall x \in S, \quad \mathbb{P}[\tau = 1, X_\tau = x] = (1 - \theta)\mu(x) \quad (1)$$

Indeed, we have for any  $x \in S$ ,

$$\begin{aligned} \mathbb{P}[\tau = 1, X_\tau = x] &= \mathbb{P}[\tau = 1, X_1 = x] \\ &= \mathbb{P}[\tau = 1 | X_1 = x] \mathbb{P}[X_1 = x] \\ &= \frac{1 - \theta}{1 - s(x)} m_1(x) \\ &= \frac{1 - \theta}{1 - s(x)} (\theta m_0(x) + (1 - \theta)\mu(x)) \\ &= \frac{1 - \theta}{1 - s(x)} \left( \theta \frac{m_0}{\mu}(x) + (1 - \theta) \right) \mu(x) \\ &= \frac{1 - \theta}{1 - s(x)} (1 - s(x)) \mu(x) \end{aligned}$$

Thus (1) is satisfied, at least for  $x \in S$  such that  $\mu(x) \neq 0$ , so that the above operations are justified. But if  $\mu(x) = 0$  and  $m_1(x) > 0$ , we have directly that  $\mathbb{P}[\tau = 1, X_1 = x] = 0$ , because  $s(x) = -\infty$ , and the equality (1) also holds in this situation. This is also trivially true if  $m_1(x) = 0$ . It follows from (1) that  $\mathbb{P}[\tau = 1] = (1 - \theta)$  and that the conditional law of  $X_1$  on  $\{\tau = 1\}$  is  $\mu$ . We also deduce that the conditional law  $\mathcal{L}(X_1 | \tau > 1)$  is equal to  $m_0$ . This observation is the key to the iterative construction, since we are led back to the initial situation, in some sense after having removed the  $\mu$  part of  $m_1$ . More rigorously, by iteration of the above procedure, at any time  $n \in \mathbb{N}$ ,  $n \geq 2$ , knowing  $X_0, X_1, \dots, X_n$  and that the value of  $\tau$  has not yet been fixed, i.e.  $\tau > n$ , we take  $\tau = n$  with probability  $(1 - \theta)/(1 - s(X_n))$ , using additional randomness as before, if necessary. The proposition will be proven if we show that for any  $n \in \mathbb{N}^*$ ,

$$\forall x \in S, \quad \mathbb{P}[\tau = n, X_n = x] = (1 - \theta)\theta^{n-1}\mu(x) \quad (2)$$

This is proved by induction on  $n$ . To facilitate this, consider the equality

$$\mathcal{L}(X_n | \tau > n) = m_0 \quad (3)$$

We have already seen that (2) and (3) are satisfied with  $n = 1$ , so let us assume they are true for any  $n \in \mathbb{N}^*$  smaller than some  $N \in \mathbb{N}^*$ , we are going to deduce them with  $n$  replaced by  $N + 1$ . Begin by computing for  $x \in S$ ,

$$\begin{aligned} \mathbb{P}[\tau = N + 1, X_{N+1} = x] &= \mathbb{P}[\tau = N + 1 | X_{N+1} = x, \tau > N] \mathbb{P}[X_{N+1} = x, \tau > N] \\ &= \frac{1 - \theta}{1 - s(x)} \mathbb{P}[X_{N+1} = x, \tau > N] \\ &= \frac{1 - \theta}{1 - s(x)} \sum_{y \in S} \mathbb{P}[X_N = y, \tau > N] P(y, x) \end{aligned}$$



where in the last line, we have used that the event  $\{\tau > N\}$  depends on  $X$  only through  $(X_m)_{0 \leq m \leq N}$ , so we can apply the Markov property. By the induction assumption, we have for any  $y \in S$ ,

$$\begin{aligned}\mathbb{P}[X_N = y, \tau > N] &= \mathbb{P}[X_N = y | \tau > N] \mathbb{P}[\tau > N] \\ &= m_0(y)(1 - \mathbb{P}[\tau \leq N]) \\ &= m_0(y)\theta^N\end{aligned}$$

where we took into account that (2) implies that  $\mathbb{P}[\tau = n] = (1 - \theta)\theta^{n-1}$  for  $n \in \llbracket 1, N \rrbracket$ . So we obtain

$$\begin{aligned}\sum_{y \in S} \mathbb{P}[X_N = y, \tau > N] P(y, x) &= \theta^N \sum_{y \in S} m_0(y) P(y, x) \\ &= \theta^N m_1(x) \\ &= \theta^N (1 - s(x)) \mu(x)\end{aligned}$$

Putting all the previous computations together, we get (2) with  $n$  replaced by  $n + 1$  (except for  $x \in S$  such that  $\mu(x) = 0$ , where (2) has to be checked directly). It remains to compute that for any  $x \in S$ ,

$$\begin{aligned}\mathbb{P}[X_{N+1} = x | \tau > N + 1] &= \frac{\mathbb{P}[X_{N+1} = x, \tau > N + 1]}{\mathbb{P}[\tau > N + 1]} \\ &= \theta^{-N-1} \mathbb{P}[X_{N+1} = x, \tau > N + 1] \\ &= \theta^{-N-1} \mathbb{P}[X_{N+1} = x, \tau \neq N + 1, \tau > N] \\ &= \theta^{-N-1} \mathbb{P}[\tau \neq N + 1 | X_{N+1} = x, \tau > N] \mathbb{P}[X_{N+1} = x, \tau > N] \\ &= \theta^{-N-1} \left(1 - \frac{1 - \theta}{1 - s(x)}\right) \mathbb{P}[X_{N+1} = x, \tau > N] \\ &= \theta^{-N-1} \frac{\theta - s(x)}{1 - s(x)} \mathbb{P}[X_{N+1} = x, \tau > N] \\ &= \theta^{-N} \frac{m_0(x)}{\mu(x)} \frac{\mathbb{P}[X_{N+1} = x, \tau > N]}{1 - s(x)}\end{aligned}$$

but as above, we have  $\mathbb{P}[X_{N+1} = x, \tau > N] = \theta^N (1 - s(x)) \mu(x)$ , so replacing it in the previous expression it appears that (3) is satisfied with  $n$  replaced by  $N + 1$ . ■

#### Remark 4

a) Contrary to the fastest strong stationary time considered in Aldous and Diaconis [1], the new strong random time constructed in Proposition 3 does not require a lot of information about the time marginal laws of the chain  $X$ . In fact only  $m_0$  and  $\theta$  are needed.

b) Notice that for a given  $m_0$ , the previous result can be applied to any  $\theta \in (0, \theta^*] \cap (0, 1)$ , with  $\theta^* = \min_{x \in S} m_0 P(x) / m_0(x)$ . The next section will help us to choose  $\theta$  in a convenient way.

c) Note that if  $x \in S$  is such that  $m_0(x) = 0$ , then  $s(x) = \theta$  (even if  $\mu(x) = 0$ , by virtue of the implicitly enforced convention that  $0 \cdot \infty = 0$ ), so for any  $n \in \mathbb{N}$ ,

$$\mathbb{P}[\tau = n + 1 | X_{n+1} = x, \tau > n] = 1$$

Thus the strong random time  $\tau$  constructed above is always bounded above by the first hitting time of the complementary set of the support of  $m_0$ . □

There is a case which is of special interest. Let us come back to the notation of the introduction, so we replace  $P$  by  $\bar{P}$  and  $S$  by  $\bar{S} = S \sqcup \{\Delta\}$ . But we now only make the first irreducibility assumption on  $\bar{P}$ . Then there exists a unique distribution  $m_0$  which is positive on  $S$  and satisfies

$$m_0 \bar{P} = \theta_0 m_0 + (1 - \theta_0) \delta_\Delta$$

where  $\theta_0 \in (0, 1)$  is the largest eigenvalue of the restriction of  $\bar{P}$  to functions vanishing on  $\Delta$  (these properties of  $\theta_0$  and  $m_0$  are insured by the Perron-Frobenius theorem, reversibility is not needed). It is indeed the renormalized positive left eigenvector associated to the eigenvalue  $\theta_0$  of  $P$  and it is called the quasi-stationary distribution of  $\bar{P}$ , because for any  $n \in \mathbb{N}$ , the restriction of  $m_0 \bar{P}^n$  to  $S$  is just  $\theta_0^n m_0$ .

Note that the function  $s$  introduced in the above proof is given by

$$\forall x \in \bar{S}, \quad s(x) = \begin{cases} \theta & , \text{ if } x = \Delta \\ -\infty & , \text{ if } x \in S \end{cases}$$

so  $\tau$  is in fact the hitting time of  $\Delta$  and we recover the well-known fact that it is distributed as a geometric law of parameter  $\theta_0$  (see for instance Section 6.5 of Chapter 3 of the book of Aldous and Fill [2]).

To profit fully from Proposition 3, it is important that no assumption is made on the respective supports of  $m_0$  and  $\mu$ .

### 3 Local equilibria

We will prove Theorem 1 here, through the use of a sequence of local equilibria, going from the initial condition  $m_0$  to the quasi-stationary measure in  $N - 1$  steps, before collapsing to  $\delta_\Delta$ .

Again we consider the discrete time framework presented in the introduction. We begin by recalling a linear algebra result of Micchelli and Willoughby [21]:

**Theorem 5** *Let  $B$  be a symmetric  $S \times S$  matrix whose entries are nonnegative and let  $b_0 \leq b_1 \leq \dots \leq b_{N-1}$  be its eigenvalues. Then for any  $l$  from  $\llbracket 0, N - 1 \rrbracket$ , the matrix  $\prod_{i \in \llbracket 0, l \rrbracket} (B - b_i)$  has nonnegative entries.*

The symmetry condition can be weakened to a reversibility condition, asking for the existence of a positive vector  $\mu := (\mu(x))_{x \in S}$  such that

$$\forall x, y \in S, \quad \mu(x)B(x, y) = \mu(y)B(y, x)$$

where we have written  $B = (B(x, y))_{x, y \in S}$ . Then let  $D$  denote the diagonal matrix with  $(\sqrt{\mu(x)})_{x \in S}$  as diagonal and define  $\tilde{B} = DBD^{-1}$ . The latter matrix is symmetric and its entries are nonnegative, furthermore it has the same eigenvalues as  $B$ . Thus for any  $l \in \llbracket 0, N - 1 \rrbracket$ , the matrix

$$\prod_{i \in \llbracket 0, l \rrbracket} (B - b_i) = D^{-1} \prod_{i \in \llbracket 0, l \rrbracket} (\tilde{B} - b_i) D$$

has also nonnegative entries.

Unless otherwise stated, from now on we assume that all the eigenvalues of  $P$  are nonnegative. We can now define the local equilibria associated to the initial condition  $m_0$ . First we consider for any  $l \in \llbracket 0, N \rrbracket$ ,

$$\tilde{\mu}_l := m_0 \prod_{i \in \llbracket 0, l-1 \rrbracket} \frac{P - \theta_{N-1-i}}{1 - \theta_{N-1-i}}$$

in particular  $\tilde{\mu}_0 = m_0$  and the Cayley-Hamilton theorem shows that  $\tilde{\mu}_N = 0$ .

From Theorem 5, applied to the reversible matrix  $P$ , we get that all these row vectors are nonnegative and thus can be considered as nonnegative measures on  $S$ . We note that their total mass is non-increasing: for any  $i \in \llbracket 0, N-1 \rrbracket$ , since  $P$  is subMarkovian, we have

$$\frac{P - \theta_{N-1-i}}{1 - \theta_{N-1-i}} \mathbb{1}_S \leq \mathbb{1}_S$$

(where  $\mathbb{1}_S$  is the indicator function of  $S$ ), so integrating this inequality with respect to  $\tilde{\mu}_i$ , we get  $\tilde{\mu}_{i+1}[S] \leq \tilde{\mu}_i[S]$ . Next we define

$$\forall i \in \llbracket 0, N-1 \rrbracket, \quad d_i := 1 - \frac{\tilde{\mu}_{i+1}[S]}{\tilde{\mu}_i[S]} \in [0, 1]$$

(in particular  $d_i = 1$  if  $\tilde{\mu}_i[S] = 0$ , since then we also have  $\tilde{\mu}_{i+1}[S] = 0$ ) and

$$\begin{aligned} I &:= \min\{i \in \llbracket 0, N-1 \rrbracket : \tilde{\mu}_{i+1}[S] = 0\} \\ &= \min\{i \in \llbracket 0, N-1 \rrbracket : d_i = 1\} \end{aligned}$$

We will show later on that  $I = N-1$ , but for the moment, we introduce for any  $i \in \llbracket 0, I \rrbracket$ , the probability measures

$$\mu_i := \left( \prod_{j \in \llbracket 0, i-1 \rrbracket} (1 - d_j)^{-1} \right) \tilde{\mu}_i$$

which are called local equilibria. Let also take the convention that  $\mu_i := \delta_\Delta$  for  $i \in \llbracket I+1, N \rrbracket$ . Their interest comes from the following observation

**Lemma 6** *For any  $i \in \llbracket 0, N-1 \rrbracket$ , we have*

$$\mu_i \bar{P} = \theta_{N-i-1} \mu_i + (1 - \theta_{N-i-1})((1 - d_i) \mu_{i+1} + d_i \delta_\Delta)$$

### Proof

By construction, we have for any  $i \in \llbracket 0, N-1 \rrbracket$ ,

$$\tilde{\mu}_i(P - \theta_{N-i-1}) = (1 - \theta_{N-i-1}) \tilde{\mu}_{i+1}$$

which can be rewritten

$$\mu_i P = \theta_{N-i-1} \mu_i + (1 - \theta_{N-i-1})(1 - d_i) \mu_{i+1}$$

For  $i \in \llbracket 0, I \rrbracket$ ,  $\mu_i$  is a probability on  $S$  and so  $\mu_i P$  is the restriction to  $S$  of the probability  $\mu_i \bar{P}$  on  $\bar{S}$ . To get the formula of the lemma, we add the missing mass, which must be at  $\Delta$ . For  $i \in \llbracket I+1, N \rrbracket$ , this formula reduces to  $\delta_\Delta = \delta_\Delta$ . ■

The proof of Theorem 1 is now quite straightforward, since it is sufficient to successively apply Proposition 3 to the local equilibria, via Lemma 6. More precisely, starting from  $m_0 = \mu_0$ , we construct a strong random time  $\tau_{N-1}$  through Proposition 3, with  $\theta = \theta_{N-1}$  and  $\mu = (1 - d_0) \mu_1 + d_0 \delta_\Delta$ . We have that  $\tau_{N-1}$  is distributed as a geometric law  $\mathcal{G}(\theta_{N-1})$  and according to Remark 4 c), with probability  $d_0$  and independently from  $\tau_{N-1}$ ,  $\tau_{N-1}$  coincides with  $T$ , the hitting time of  $\Delta$ , and with probability  $1 - d_0$ ,  $X_{\tau_{N-1}}$  is distributed according to the second local equilibrium  $\mu_1$ . Let us define the shifted chain  $\tilde{X}^{(1)} := (\tilde{X}_n^{(1)})_{n \in \mathbb{N}} := (X_{\tau_{N-1}+n})_{n \in \mathbb{N}}$ . Due to the fact that  $\tau_{N-1}$  is a strong random time,  $\tilde{X}^{(1)}$  is independent from  $\tau_{N-1}$  and is a Markov chain whose transition

matrix is  $\bar{P}$  and with  $(1 - d_1)\mu_1 + d_1\delta_\Delta$  as initial distribution. Let us consider  $X^{(1)}$  which is just  $\tilde{X}^{(1)}$  conditioned by  $X_0^{(1)} \neq \Delta$ . It is a Markov chain admitting  $\bar{P}$  as transition matrix and  $\mu_1$  as initial distribution. Apply Proposition 3, with  $m_0 = \mu_1$ ,  $\theta = \theta_{N-2}$  and  $\mu = (1 - d_1)\mu_2 + d_1\delta_\Delta$ , to construct a strong random time  $\tau_{N-2}$  for  $X^{(1)}$ , which is distributed as  $\mathcal{G}(\theta_{N-2})$  and is independent of  $\tau_{N-1}$ .

The proof goes on by an obvious iteration and in the end (the procedure has to stop, because  $I \leq N - 1$ , due to the fact that  $\tilde{\mu}_N = 0$ ), we get Theorem 1 with

$$\forall i \in \llbracket 0, N - 1 \rrbracket, \quad a_{N-1-i} := d_i \prod_{j \in \llbracket 0, i-1 \rrbracket} (1 - d_j) \quad (4)$$

Indeed, one can define a random variable  $A$  with distribution  $a = (a_0, a_1, \dots, a_{N-1})$  (see the discussion before Theorem 1) in the following way: let us denote  $\tilde{X}^{(1)}, \tilde{X}^{(2)}, \dots$ , the shifted Markov chains constructed iteratively through the previous procedure, then take

$$A := N - \min\{i \in \llbracket 1, N \rrbracket : \tilde{X}_0^{(i)} = \Delta\}$$

Then for any  $i \in \llbracket 1, N \rrbracket$ , the event  $\{A = i\}$  is independent of  $\tau_{N-1}, \dots, \tau_{N-i}$  (which are also independent among themselves). This is sufficient to insure that  $T = \sum_{j \in \llbracket N-A, N-1 \rrbracket} \tau_j$  is distributed according to  $\mathcal{M}_{\mathcal{G}}(a, \theta)$ .

The only point which remains to be proven is that  $a_0 > 0$ . This is a consequence of the well-known fact that for any initial distribution  $m_0$  on  $S$ ,

$$\lim_{n \rightarrow \infty} \frac{1}{n} \ln(\mathbb{P}[T > n]) = -\ln(\theta_0) \quad (5)$$

(indeed, one can be brought back to the case where the initial distribution is the quasi-stationary probability by noticing that the quantities  $\lim_{n \rightarrow \infty} \frac{1}{n} \ln(\mathbb{P}_x[T > n])$  exist and do not depend on the starting point  $x \in S$  by the irreducibility assumption). But if  $J = \min\{j \in \llbracket 0, N - 1 \rrbracket : a_j > 0\}$ , and if  $T$  is distributed according to  $\mathcal{M}_{\mathcal{G}}(a, \theta)$ , then we have that

$$\forall 0 \leq u < \frac{1}{\theta_J}, \quad \mathbb{E}[u^T] < +\infty$$

so the fact that  $\theta_0 > \theta_1$  shows that  $J = 0$ . Furthermore, since

$$\begin{aligned} a_0 &= d_{N-1} \prod_{j \in \llbracket 0, N-2 \rrbracket} (1 - d_j) \\ &= \prod_{j \in \llbracket 0, N-2 \rrbracket} (1 - d_j) \end{aligned}$$

we also get that  $d_j < 1$  for all  $j \in \llbracket 0, N - 2 \rrbracket$ , so  $I = N - 1$ . Thus  $\mu_{N-1}$  is a probability on  $S$ , and since Lemma 6 shows that

$$\mu_{N-1} \bar{P} = \theta_0 \mu_{N-1} + (1 - \theta_0) \delta_\Delta$$

it follows that  $\mu_{N-1}$  is the quasi-stationary probability associated to  $\bar{P}$  (recall its definition given in Remark 4c).

**Remark 7** It is the sequence of local equilibria  $(\mu_i)_{i \in \llbracket 0, N-1 \rrbracket}$  which has a meaning, not the notion of local equilibrium by itself. Indeed, we can go a little further: let say that a finite sequence of probabilities on  $S$ ,  $(\mu_i)_{i \in \llbracket M, N-1 \rrbracket}$ , with  $M \in \llbracket 0, N - 1 \rrbracket$ , is an  $M$ -sequence of local equilibria going from  $\mu_M$  to  $\delta_\Delta$ , if the relations of Lemma 6 are satisfied for  $i \in \llbracket M, N - 1 \rrbracket$ , with a sequence  $(d_i)_{i \in \llbracket M, N-1 \rrbracket}$  of numbers from  $[0, 1]$  and with  $\mu_N = \delta_\Delta$ . The above proof shows that if the initial distribution of  $X$  is  $\mu_M$ , then the law of  $T$  is a mixture of the distributions  $\mathcal{G}(\theta_i, \theta_{i+1}, \dots, \theta_{N-1-M})$ ,

for  $i \in \llbracket 0, N-1-M \rrbracket$ , which gives positive weight to  $\mathcal{G}(\theta_0, \theta_1, \dots, \theta_{N-1-M})$ . This is even true under the weakened assumption that  $\theta_{N-1-M} \geq 0$ , so the eigenvalues  $\theta_{N-M}, \theta_{N-M+1}, \dots, \theta_{N-1}$  can be negative.

In particular if  $M = N-1$ ,  $\mu_{N-1}$  must be the quasi-stationary probability and we know by the Perron-Frobenius theorem that  $\theta_0$  is nonnegative. So we recover the fact that if the initial distribution of  $X$  is the quasi-stationary probability, then the distribution of  $T$  is a geometric law of parameter  $\theta_0$ . Note that the reversibility hypothesis can be dispensed with in this special case, since we don't need Theorem 5 to check that  $\mu_N = \delta_\Delta$  is nonnegative.

It also appears that to write the law of  $T$  in the form  $\mathcal{M}_\mathcal{G}(\theta, a)$  may not be very clever and that it is better to introduce for  $M \in \llbracket 0, N-1 \rrbracket$ , the distributions

$$\mathcal{M}_\mathcal{G}(M, \theta, a) := \sum_{0 \leq i \leq N-1-M} a_i \mathcal{G}(\theta_i, \theta_{i+1}, \dots, \theta_{N-1-M})$$

where  $a$  is a probability on  $\llbracket 0, N-1-M \rrbracket$ . Now the law of  $T$  may not be uniquely represented as  $\mathcal{M}_\mathcal{G}(M, \theta, a)$ . Indeed, we may have  $\mathcal{M}_\mathcal{G}(M, \theta, a) = \mathcal{M}_\mathcal{G}(M', \theta, a')$  with  $M \neq M'$  and  $a \neq a'$ . For instance, taking into account the previous example,  $\mathcal{M}_\mathcal{G}(N-1, \theta, 1) = \mathcal{G}(\theta_0)$ , for any  $M \in \llbracket 0, N-2 \rrbracket$ , we can find a probability  $a_M$  on  $\llbracket 0, N-1-M \rrbracket$  such that  $\mathcal{G}(\theta_0) = \mathcal{M}_\mathcal{G}(M, \theta, a_M)$ . To simplify the expression for the law of  $T$ , one has to try to find the largest  $M$  such that there exists an  $M$ -sequence of local equilibria going from  $m_0$  to  $\delta_\Delta$ .

Note also that for given  $M \in \llbracket 0, N-1 \rrbracket$  and a nonnegative  $\theta = (\theta_0, \dots, \theta_{N-1})$ , the mapping  $a = (a_0, \dots, a_{N-1-M}) \mapsto \mathcal{M}_\mathcal{G}(\theta, a)$  is one to one. This can be proven by considering the corresponding moment generating functions (or rather its meromorphic extension to  $\mathbb{C}$ ).

Finally notice that from a complexity point of view, once the eigenvalues are known,  $(\mu_i)_{i \in \llbracket 0, N-1 \rrbracket}$ ,  $(d_i)_{i \in \llbracket 0, N-1 \rrbracket}$  and  $(a_i)_{i \in \llbracket 0, N-1 \rrbracket}$  are not difficult to obtain via Lemma 6 and (4), indeed at most  $N$  iterations of  $\bar{P}$  are necessary. To obtain the largest  $M$  such that the law of  $T$  can be represented in the form  $\mathcal{M}_\mathcal{G}(M, \theta, a)$ , at most  $(N-1)N/2$  steps are needed, since we have first to try with  $M = N-1$ , next  $M = N-2$  etc. to verify if we can associate to  $m_0$  an  $M$ -sequence of local equilibria through Lemma 6.

□

## 4 Intertwining

Our goal here is to construct an intertwining relation between  $X$  and a Markov chain of a special kind. This is indeed the main result of this paper, as it gives a probabilistic interpretation of Theorem 1.

More precisely, let  $\bar{P}$  be an absorbing transition matrix on  $\bar{S}$ , as in the introduction, and let  $\theta_0 > \theta_1 \geq \dots \geq \theta_{N-1} \geq 0$  be its eigenvalues, assumed to be nonnegative. For a given initial distribution  $m_0$ , consider the associated sequences of local equilibria  $(\mu_i)_{i \in \llbracket 0, N-1 \rrbracket}$  and of real numbers from  $[0, 1]$ ,  $(d_i)_{i \in \llbracket 0, N-1 \rrbracket}$ , as in Lemma 6. Recall that for  $i \in \llbracket 0, N-2 \rrbracket$ ,  $d_i < 1$  and that  $d_{N-1} = 1$ .

Next introduce an absorbing Markov chain  $Y := (Y_n)_{n \in \mathbb{N}}$  on  $\bar{E} := \llbracket 0, N-1 \rrbracket \sqcup \{\Delta\}$ , starting from 0 and with transition matrix  $\bar{Q}$  given by

$$\forall i, j \in \bar{E}, \quad \bar{Q}(i, j) := \begin{cases} \theta_{N-1-i} & , \text{ if } i = j \in \llbracket 0, N-1 \rrbracket \\ d_i(1 - \theta_{N-1-i}) & , \text{ if } i \in \llbracket 0, N-1 \rrbracket \text{ and } j = \Delta \\ (1 - d_i)(1 - \theta_{N-1-i}) & , \text{ if } i \in \llbracket 0, N-2 \rrbracket \text{ and } j = i+1 \\ 1 & , \text{ if } i = j = \Delta \\ 0 & , \text{ otherwise} \end{cases} \quad (6)$$

(for  $X$  as well as for  $Y$ , we denote  $\Delta$  the absorbing point, this should not lead to confusion).

The behaviour of the Markov chain  $Y$  is easy to understand: it tries to go to the next neighbour on the right, or stay at the same place, except that it can also be killed (i.e. go to the absorbing point) and the respective probabilities for these actions are  $(1-d_i)(1-\theta_{N-1-i})$ ,  $\theta_{N-1-i}$  and  $d_i(1-\theta_{N-1-i})$  when the current position is  $i \in \llbracket 0, N-2 \rrbracket$ . This is also true if  $i = N-1$  since  $d_{N-1} = 1$ , so from there the Markov chain either stays at the same place, with probability  $\theta_0$ , or is killed, with probability  $1-\theta_0$ . But for  $i \in \llbracket 0, N-2 \rrbracket$ , the probability to go to the right is positive. With this interpretation, for  $i \in \llbracket 0, N-1 \rrbracket$ ,  $d_i$  can be called the chance of death immediately after moving from the  $i^{\text{th}}$  local equilibrium.

Let us furthermore denote by  $\Lambda$  the  $|\bar{E}| \times |\bar{S}|$  matrix defined by

$$\forall i \in \bar{E}, \forall x \in \bar{S}, \quad \Lambda(i, x) := \begin{cases} \mu_i(x) & , \text{ if } i \in \llbracket 0, N-1 \rrbracket \\ \delta_\Delta(x) & , \text{ if } i = \Delta \end{cases} \quad (7)$$

Its interest comes from the following intertwining relation

**Lemma 8** *For  $\bar{Q}$  and  $\Lambda$  defined at (6) and (7), we have the intertwining relation*

$$\Lambda \bar{P} = \bar{Q} \Lambda$$

**Proof**

This is just a rewriting of Lemma 6: fix some  $i \in \llbracket 0, N-1 \rrbracket$  and  $x \in \bar{S}$ . On one hand we have

$$\begin{aligned} \Lambda \bar{P}(i, x) &= \sum_{y \in \bar{S}} \Lambda(i, y) \bar{P}(y, x) \\ &= \mu_i \bar{P}(x) \end{aligned}$$

and on the other hand,

$$\begin{aligned} \bar{Q} \Lambda(i, x) &= \theta_{N-1-i} \Lambda(i, x) + (1-d_i)(1-\theta_{N-1-i}) \Lambda(i+1, x) + d_i(1-\theta_{N-1-i}) \Lambda(\Delta, x) \\ &= (\theta_{N-1-i} \mu_i + (1-\theta_{N-1-i})((1-d_i) \mu_{i+1} + d_i \delta_\Delta))(x) \end{aligned}$$

from where we get the equality  $\Lambda \bar{P}(i, x) = \bar{Q} \Lambda(i, x)$ .

Consider finally the case where  $i = \Delta$ . Then the previous equality reads  $\delta_\Delta(x) = \delta_\Delta(x)$ . ■

This intertwining relation and the fact that we have for the initial laws

$$\mathcal{L}(Y_0) \Lambda = m_0 = \mathcal{L}(X_0)$$

enables use of a result of Diaconis and Fill [5] asserting we can intertwine  $X$  in  $Y$  in the following sense: there exists a coupling of  $X$  and  $Y$  such that the chain  $(X_n, Y_n)_{n \in \mathbb{N}}$  is Markovian and such that for any time  $n \in \mathbb{N}$ , we have for the conditional law of  $X_n$  knowing  $Y_0, Y_1, \dots, Y_n$ ,

$$\mathcal{L}(X_n | Y_0, Y_1, \dots, Y_n) = \Lambda(Y_n, \cdot)$$

In some sense  $Y$  indicates which local equilibrium has been reached by  $X$  and in this respect, it is interesting to know that if a trajectory  $(X_n)_{n \in \mathbb{N}}$  is given, it is possible to construct  $(Y_n)_{n \in \mathbb{N}}$  from it in a progressive way, using also independent randomness (for more details, see Diaconis and Fill [5]).

Under the above coupling, let us denote

$$\begin{aligned} T_X &:= \inf\{n \in \mathbb{N} : X_n = \Delta\} \\ T_Y &:= \inf\{n \in \mathbb{N} : Y_n = \Delta\} \end{aligned}$$

It is clear that  $T_X$  has the same law as  $T$  and that  $T_Y$  is distributed according to  $\mathcal{M}_G(\theta, a)$ , with the probability  $a$  given by (4). So the next result gives another proof of Theorem 1:

**Lemma 9** *We have a.s. that  $T_X = T_Y$ .*

**Proof**

Let  $n \in \mathbb{N}$  be given, we have that

$$\begin{aligned} \mathbb{P}[T_X \leq n, T_Y \leq n] &= \mathbb{E}[\delta_\Delta(X_n)\delta_\Delta(Y_n)] \\ &= \mathbb{E}[\mathbb{E}[\delta_\Delta(X_n)|Y_0, \dots, Y_n]\delta_\Delta(Y_n)] \\ &= \mathbb{E}[\Lambda(Y_n, \Delta)\delta_\Delta(Y_n)] \\ &= \Lambda(\Delta, \Delta)\mathbb{E}[\delta_\Delta(Y_n)] \\ &= \mathbb{P}[T_Y \leq n] \end{aligned}$$

so it follows that  $\{T_X \leq n\}$  is a.s. included in  $\{T_Y \leq n\}$ . Since this is true for any  $n \in \mathbb{N}$ , we get that  $T_X \leq T_Y$  a.s.

But for any  $n \in \mathbb{N}$ , we compute in a similar way that

$$\begin{aligned} \mathbb{P}[T_X \leq n, T_Y > n] &= \mathbb{E}[\Lambda(Y_n, \Delta)\mathbb{1}_{[0, N-1]}(Y_n)] \\ &= 0 \end{aligned}$$

and as a consequence  $T_X = T_Y$  a.s. ■

Thus for any absorbing Markov chain satisfying the hypotheses of the introduction, the absorption time is distributed according to a law of the form  $\mathcal{M}_G(\theta, a)$ , with  $\theta \in \Theta$  and  $a \in \mathcal{A}$ , where

$$\begin{aligned} \Theta &:= \{\theta = (\theta_i)_{i \in \llbracket 0, N-1 \rrbracket} \in [0, 1]^N : \theta_0 > \theta_1 \geq \theta_2 \geq \dots \geq \theta_{N-1}\} \\ \mathcal{A} &:= \{a = (a_i)_{i \in \llbracket 0, N-1 \rrbracket} \in \mathcal{P}(\llbracket 0, N-1 \rrbracket) : a_0 > 0\} \end{aligned}$$

where from now on  $\mathcal{P}(E)$  will denote the set of probability measures on  $E$ , for any finite set  $E$ . Conversely, we can wonder if for an a priori given  $\theta \in \Theta$  and  $a \in \mathcal{A}$ , one can find an absorbing Markov chain satisfying the assumptions of the introduction and such that the law of its absorption time is distributed according to  $\mathcal{M}_G(\theta, a)$ . Then the set of laws of such absorption times would exactly be  $\mathcal{M}_G := \{\mathcal{M}_G(\theta, a) : \theta \in \Theta, a \in \mathcal{A}\}$ . We do not conjecture that this assertion is true, because we believe more restrictions should be put on the admissible pairs  $(\theta, a)$ , for instance maybe that for any  $i \in \llbracket 1, N-1 \rrbracket$ ,  $a_i > 0 \Rightarrow a_{i-1} > 0$ .

Nevertheless, the above chain  $Y$  can be used to get a result in this direction. Let  $\bar{\mathcal{K}}$  be the closure of the set of Markov transition matrices  $\bar{P}$  on  $\bar{S}$  whose unique recurrence class is  $\{\Delta\}$  and such that the restriction  $P$  of  $\bar{P}$  to  $S \times S$  is irreducible, reversible and admits nonnegative eigenvalues. Note that the elements of  $\bar{\mathcal{K}}$  are still transition matrices absorbing at  $\Delta$ . Here we are interested in  $\hat{\mathcal{K}}$ , the set of matrices  $\bar{P}$  from  $\bar{\mathcal{K}}$  such that  $\{\infty\}$  is the unique recurrence class of  $\bar{P}$ . If  $m_0 \in \mathcal{P}(S)$  and  $\bar{P} \in \hat{\mathcal{K}}$  are given, we can associate to them a Markov chain, starting from  $m_0$  with transition matrix  $\bar{P}$ . Let us write  $\mathcal{L}(m_0, \bar{P})$  for the law on  $\mathbb{N}^*$  of the corresponding absorption time. We also need to introduce

$$\widehat{\mathcal{M}}_G := \{\mathcal{M}_G(\theta, a) : \theta \in \hat{\Theta}, a \in \mathcal{P}(\llbracket 0, N-1 \rrbracket)\}$$

with

$$\hat{\Theta} := \{\theta = (\theta_i)_{i \in \llbracket 0, N-1 \rrbracket} \in [0, 1]^N : \theta_0 \geq \theta_1 \geq \theta_2 \geq \dots \geq \theta_{N-1}\}$$

**Proposition 10** *The set of distributions  $\mathcal{L}(m_0, \bar{P})$ , when  $m_0$  runs through  $\mathcal{P}(S)$  and  $\bar{P}$  runs through  $\hat{\mathcal{K}}$ , coincides with  $\widehat{\mathcal{M}}_G$ .*

## Proof

We begin with the inclusion

$$\{\mathcal{L}(m_0, \bar{P}) : m_0 \in \mathcal{P}(S), \bar{P} \in \widehat{\mathcal{K}}\} \subset \widehat{\mathcal{M}}_{\mathcal{G}} \quad (8)$$

Let  $\bar{P} \in \widehat{\mathcal{K}}$  be given and  $(\bar{P}^{(r)})_{r \in \mathbb{N}}$  be a sequence of transition matrices as in the introduction and converging to  $\bar{P}$ . For any fixed  $r \in \mathbb{N}$ , we consider  $1 > \theta_0^{(r)} > \theta_1^{(r)} \geq \dots \geq \theta_{N-1}^{(r)} \geq 0$  the eigenvalues of the restriction  $P^{(r)}$  of  $\bar{P}^{(r)}$  to  $S \times S$ . Let us also fix an initial distribution  $m_0 \in \mathcal{P}(S)$ . Lemma 6 enables us to associate to  $m_0$  and  $\bar{P}^{(r)}$  a sequence of local equilibria  $(\mu_i^{(r)})_{i \in \llbracket 0, N-1 \rrbracket}$  from  $\mathcal{P}(S)$  and a sequence of real numbers from  $[0, 1]$ ,  $(d_i^{(r)})_{i \in \llbracket 0, N-2 \rrbracket}$ . As in the beginning of this section, we construct from them two matrices  $\bar{Q}^{(r)}$  and  $\Lambda^{(r)}$  so that

$$\Lambda^{(r)} \bar{P}^{(r)} = \bar{Q}^{(r)} \Lambda^{(r)} \quad (9)$$

Up to the extraction of a subsequence, we can assume by standard compactness arguments, that there exist  $1 \geq \theta_0 \geq \theta_1 \geq \dots \geq \theta_{N-1} \geq 0$ , a sequence  $(\mu_i)_{i \in \llbracket 0, N-1 \rrbracket}$  of probabilities from  $\mathcal{P}(S)$  and a sequence of real numbers from  $[0, 1]$ ,  $(d_i)_{i \in \llbracket 0, N-2 \rrbracket}$  such that

$$\begin{aligned} \forall i \in \llbracket 0, N-1 \rrbracket, \quad \lim_{r \rightarrow \infty} \theta_i^{(r)} &= \theta_i \\ \forall i \in \llbracket 0, N-1 \rrbracket, \quad \lim_{r \rightarrow \infty} \mu_i^{(r)} &= \mu_i \\ \forall i \in \llbracket 0, N-2 \rrbracket, \quad \lim_{r \rightarrow \infty} d_i^{(r)} &= d_i \end{aligned}$$

Thus if we construct  $\bar{Q}$  and  $\Lambda$  from  $(\theta_i)_{i \in \llbracket 0, N-1 \rrbracket}$ ,  $(d_i)_{i \in \llbracket 0, N-2 \rrbracket}$  and  $(\mu_i)_{i \in \llbracket 0, N-1 \rrbracket}$  as above, we get, by passing to the limit in (9)

$$\Lambda \bar{P} = \bar{Q} \Lambda$$

We note that the first line of  $\Lambda$  is  $m_0$  (since this is also true for  $\Lambda^{(r)}$ , for all  $r \in \mathbb{N}$ ), so using again a result of Diaconis and Fill [5], we can intertwine, as above, a Markov chain  $X$  starting from  $m_0$  and admitting  $\bar{P}$  as transition matrix with a Markov chain  $Y$  starting from  $\delta_0$  and admitting  $\bar{Q}$  as transition matrix. Lemma 9 is still valid and we deduce that  $\mathcal{L}(m_0, \bar{P})$  is equally the distribution of the hitting time  $T_Y$  of  $\Delta$  by  $Y$ . But the latter can now be a.s. infinite and one easily checks that it is a.s. finite if and only if  $\theta_0 < 1$ . To deduce this property, we have to take into account that the only recurrence class of  $\bar{P}$  is  $\{\Delta\}$ . Indeed, this latter assumption implies that 1 cannot be an eigenvalue of  $P$ , the restriction of  $\bar{P}$  to  $S \times S$ . But by a standard finite dimensional perturbation result (see for instance the book [16] of Kato), we have that the eigenvalues of the matrices  $P^{(r)}$  converge to the eigenvalues of  $P$ , so in particular  $\theta_0$  is the largest eigenvalue of  $P$  and thus  $\theta_0 < 1$ . Now it is clear from the structure of  $\bar{Q}$  that the distribution of  $T_Y$  belongs to  $\widehat{\mathcal{M}}_{\mathcal{G}}$  and this observation concludes the proof of (8).

We now come to the reverse inclusion

$$\{\mathcal{L}(m_0, \bar{P}) : m_0 \in \mathcal{P}(S), \bar{P} \in \widehat{\mathcal{K}}\} \supset \widehat{\mathcal{M}}_{\mathcal{G}} \quad (10)$$

Let  $\theta := (\theta_i)_{i \in \llbracket 0, N-1 \rrbracket} \in \widehat{\Theta}$  and  $a \in \mathcal{P}(\llbracket 0, N-1 \rrbracket)$  be given. We have first to verify that we can find a sequence  $(d_i^{(r)})_{i \in \llbracket 0, N-2 \rrbracket}$  of real numbers from  $[0, 1]$ , such that

$$\forall i \in \llbracket 0, N-1 \rrbracket, \quad a_{N-1-i} := d_i \prod_{j \in \llbracket 0, i-1 \rrbracket} (1 - d_j)$$

Seeing  $a$  as a probability on  $\llbracket 0, N-1 \rrbracket$ , let

$$I := \max\{i \in \llbracket 0, N-1 \rrbracket : a(\llbracket 0, i-1 \rrbracket) = 0\}$$



so  $I = 0$  is equivalent to  $a_0 > 0$ . Next we define

$$\forall i \in \llbracket 0, N-2 \rrbracket, \quad d_i := \begin{cases} \frac{a_{N-1-i}}{a(\llbracket 0, N-1-i \rrbracket)} & , \text{ if } i \leq N-I-1 \\ 1 & , \text{ otherwise} \end{cases} \quad (11)$$

and we immediately check that the above relation between  $a$  and  $(d_i)_{i \in \llbracket 0, N-2 \rrbracket}$  is satisfied. Now we have at our disposition all the ingredients to construct the matrix  $\bar{Q}$  given in (6). Furthermore, it is quite obvious that  $\mathcal{L}(\delta_0, \bar{Q}) = \mathcal{M}_G(\theta, a)$ . So to end the proof of (10), it remains to check that  $\bar{Q} \in \hat{\mathcal{K}}$ . First we remark that since  $1 > \theta_0 \geq \theta_1 \geq \dots \geq \theta_{N-I-1} \geq 0$  and  $d_{N-I-1} = 1$ , the path  $(0, 1, 2, \dots, N-I-1, \Delta)$  is admissible for  $\bar{Q}$ , as well as the one-step paths  $(i, \Delta)$ , for  $i \in \llbracket N-I, N-1 \rrbracket$ , because  $1 > \theta_i \geq 0$  and  $d_i = 1$ . So  $\{\Delta\}$  is the unique recurrence class for  $\bar{Q}$ . Next we define, for  $\epsilon \in (0, 1/2)$ , a new absorbing transition matrix  $\bar{Q}_\epsilon$  by taking

$$\forall i, j \in \bar{E}, \quad \bar{Q}_\epsilon(i, j) := \begin{cases} \epsilon & , \text{ if } i \in \llbracket 1, N-1 \rrbracket \text{ and } j = i-1 \\ \epsilon & , \text{ if } i \in \llbracket N-I-1, N-2 \rrbracket \text{ and } j = i+1 \\ (1-\epsilon)\theta_{N-1-i} & , \text{ if } i = j \in \llbracket 0, N-I-1 \rrbracket \\ (1-2\epsilon)\theta_{N-1-i} & , \text{ if } i = j \in \llbracket N-I, N-1 \rrbracket \\ (1-\epsilon)d_i(1-\theta_{N-1-i}) & , \text{ if } i \in \llbracket 0, N-I-1 \rrbracket \text{ and } j = \Delta \\ (1-2\epsilon)(1-\theta_{N-1-i}) & , \text{ if } i \in \llbracket N-I, N-1 \rrbracket \text{ and } j = \Delta \\ (1-\epsilon)(1-d_i)(1-\theta_{N-1-i}) & , \text{ if } i \in \llbracket 0, N-I-2 \rrbracket \text{ and } j = i+1 \\ 1 & , \text{ if } i = j = \Delta \\ 0 & , \text{ otherwise} \end{cases}$$

As usual, let  $Q_\epsilon$  be the restriction of  $\bar{Q}_\epsilon$  to  $\llbracket 0, N-1 \rrbracket^2$ . It is an irreducible subMarkovian matrix of birth and death type and thus  $Q_\epsilon$  is reversible. Let us denote by

$$1 > \theta_0(\epsilon) > \theta_1(\epsilon) > \theta_2(\epsilon) > \dots > \theta_{N-1}(\epsilon) > -1$$

the eigenvalues of  $Q_\epsilon$  (they are well-known to be all distinct in the context of irreducible subMarkovian birth and death chains). It may happen that some of them are negative. This is for instance the case, as soon as  $N \geq 1$ , if all the  $\theta_i$ , for  $i \in \llbracket 0, N-1 \rrbracket$ , are zero and if  $a = \delta_{N-1}$ , in which case  $Q_\epsilon = \epsilon(\delta_1(|i-j|))_{i,j \in \llbracket 0, N-1 \rrbracket}$  (it has some negative eigenvalue, because it is symmetric, not null and its trace is zero). So let us rather consider

$$\hat{Q}_\epsilon := \alpha(\epsilon)\text{Id} + (1-\alpha(\epsilon))\bar{Q}_\epsilon$$

with  $\alpha(\epsilon) := \max(0, -\theta_{N-1}(\epsilon)) \in [0, 1)$ . Since  $\lim_{\epsilon \rightarrow 0+} \bar{Q}_\epsilon = \bar{Q}$ , we also have

$$\lim_{\epsilon \rightarrow 0+} \theta_{N-1}(\epsilon) = \theta_{N-1}$$

and by consequence  $\lim_{\epsilon \rightarrow 0+} \alpha(\epsilon) = 0$ . Thus we get

$$\lim_{\epsilon \rightarrow 0+} \hat{Q}_\epsilon = \bar{Q}$$

So finally  $\bar{Q} \in \bar{\mathcal{K}}$ , because all the transition matrices  $\bar{Q}_\epsilon$ , for  $\epsilon \in (0, 1/2)$ , satisfy the assumption of the introduction. ■

## Remarks 11

1) The end of the above proof shows that the matrix  $\bar{Q}$  constructed from  $\bar{P}$  in (6) belongs to  $\hat{\mathcal{K}}$ . Clearly,  $Q$  is not reversible, since moves to the left are not permitted on  $\llbracket 1, N-1 \rrbracket$ . Also notice that since  $Q$ , the restriction of  $\bar{Q}$  to  $\llbracket 0, N-1 \rrbracket^2$  (respectively  $\bar{Q}$ ) is upper triangular, its eigenvalues

are immediate to compute and are given by the entries of the diagonal of  $Q$  (resp.  $\bar{Q}$ ). But if some of these eigenvalues are equal,  $Q$  (resp.  $\bar{Q}$ ) is not necessary diagonalizable, e.g. this is the case for the following matrix

$$\bar{Q} := \begin{pmatrix} \theta & 1 & 0 \\ 0 & \theta & 1 \\ 0 & 0 & 1 \end{pmatrix}$$

for any  $\theta \in [0, 1)$ .

2) Let  $(\bar{P}^{(r)})_{r \in \mathbb{N}}$  be a sequence from  $\hat{\mathcal{K}}$  converging to  $\bar{P} \in \hat{\mathcal{K}}$  and  $(m_0^{(r)})_{r \in \mathbb{N}}$  be a sequence from  $\mathcal{P}(S)$  converging to  $m_0 \in \mathcal{P}(S)$ . Then for large  $r$ ,  $\mathcal{L}(m_0^{(r)}, \bar{P}^{(r)})$  is weakly convergent to  $\mathcal{L}(m_0, \bar{P})$ , because this is equivalent to

$$\forall k \in \mathbb{N}^*, \quad \lim_{r \rightarrow \infty} \mathcal{L}(m_0^{(r)}, \bar{P}^{(r)})[k] = \mathcal{L}(m_0, \bar{P})[k]$$

and to get these limits, just write  $\mathcal{L}(m_0^{(r)}, \bar{P}^{(r)})[k]$  as a finite sum of products of  $k + 1$  entries from  $m_0^{(r)}$  or  $\bar{P}^{(r)}$ .

Assume that there exists  $m_0 \in \mathcal{P}(S)$  such that  $\mathcal{L}(m_0, \bar{P}) = \mathcal{M}_G(\theta, a)$  with  $a_0 > 0$ . Then by considering the moment generating functions of the  $\mathcal{L}(m_0, \bar{P}^{(r)})$ , for  $r \in \mathbb{N}$ , and of  $\mathcal{L}(m_0, \bar{P})$ , we can deduce the convergence for large  $r \in \mathbb{N}$  of the eigenvalues of  $\bar{P}^{(r)}$  to the eigenvalues of  $\bar{P}$ . This gives a probabilistic alternative to the reference to classical perturbation theory in the proof of Proposition 10.

3) Let  $\tilde{\mathcal{K}}$  denote the set of  $\bar{P} \in \hat{\mathcal{K}}$  which are weakly irreducible, in the sense that we cannot decompose  $S$  into a non trivial partition  $S = S_1 \sqcup S_2$  such that the restrictions of  $\bar{P}$  to  $S_1 \times S_2$  and to  $S_2 \times S_1$  are null. We also introduce

$$\tilde{\mathcal{M}}_G := \{ \mathcal{M}_G(\theta, a) : \theta \in \hat{\Theta}, a \in \mathcal{P}([0, N - 1]) \text{ with } a_0 > 0 \}$$

It seems possible that

$$\{ \mathcal{L}(m_0, \bar{P}) : m_0 \in \mathcal{P}(S), \bar{P} \in \tilde{\mathcal{K}} \} = \tilde{\mathcal{M}}_G$$

□

In the same spirit, note that due to the example of Botta, Harris and Marchal [3] alluded to in the introduction, Theorems 1 and 2 are no longer true when the reversibility condition is replaced by the assumption that all the eigenvalues of  $P$  are nonnegative. It follows that there are such subMarkovian matrices  $P$  that cannot be approached by reversible subMarkovian matrices (but maybe this is true for real-diagonalizable subMarkovian matrices). To understand better this phenomenon, maybe it would be interesting to investigate carefully the case of skip-free chains considered by Fill [12].

## 5 Continuous time

We now return to the continuous-time framework presented in the introduction. We will not develop a specific continuous time approach, instead we extend the previous results by considering discrete time skeletons of the underlying semi-group.

More precisely, as already mentioned in the introduction, under the hypotheses made in the continuous-time setting, for any  $\delta > 0$ , the absorbing transition matrix  $\bar{P}_\delta = \exp(\delta \bar{L})$  satisfies the previous discrete time assumptions:  $\{\triangle\}$  is the unique recurrent class of  $\bar{P}_\delta$  and its restriction  $P_\delta$  to  $S \times S$ , which coincides with  $\exp(\delta L)$ , is irreducible and reversible with respect to the same positive

probability  $\pi$  on  $S$ . Let  $X = (X_t)_{t \geq 0}$  be a Markov process with cadlag trajectories, starting from  $m_0$  with generator  $\bar{L}$ . Then  $X^{(\delta)} := (X_{\delta n})_{n \in \mathbb{N}}$  is a Markov chain starting from  $m_0$  with transition matrix  $\bar{P}_\delta$ . Recall that we are interested in  $T$ , the absorption time of  $X$ , and let us denote

$$T^{(\delta)} := \inf\{n \in \mathbb{N} : X_{\delta n} = \triangle\}$$

the absorption time of  $X^{(\delta)}$ .

The key to Theorem 2 is the simple observation that

**Lemma 12** *As  $\delta$  goes to  $0_+$ ,  $\delta T^{(\delta)}$  converges a.s. to  $T$ , in particular in the weak sense,*

$$\lim_{\delta \rightarrow 0_+} \mathcal{L}(\delta T^{(\delta)}) = \mathcal{L}(T)$$

### Proof

Since the trajectories of  $X$  are cadlag, for any  $s \geq 0$ ,  $X_{T+s} = \triangle$ , at least outside the negligible event where  $T = +\infty$ , so

$$T \leq \delta T^{(\delta)} \leq T + \delta$$

■

Let  $\theta^{(\delta)} = (\theta_i^{(\delta)})_{i \in \llbracket 0, N-1 \rrbracket}$  be the eigenvalues of  $P_\delta$ , with multiplicities and in nonincreasing order. Since  $\theta_i^{(\delta)} = \exp(-\delta \lambda_i)$  for all  $i \in \llbracket 0, N-1 \rrbracket$ , they are positive, so we can apply Theorem 1 to get a probability  $a^{(\delta)}$  on  $\llbracket 0, N-1 \rrbracket$  such that

$$\mathcal{L}(T^{(\delta)}) = \mathcal{M}_G(\theta^{(\delta)}, a^{(\delta)}) \quad (12)$$

By compactness of  $\mathcal{P}(\llbracket 0, N-1 \rrbracket)$ , a probability  $a$  can be found on  $\llbracket 0, N-1 \rrbracket$  and a decreasing sequence  $(\delta^{(r)})_{r \in \mathbb{N}}$  of positive real numbers converging to 0, such that  $\lim_{r \rightarrow \infty} a^{(\delta^{(r)})} = a$ . In conjunction with the above lemma. The next result enables us to conclude the proof of Theorem 2, except for the assertion that  $a_0 > 0$ .

**Lemma 13** *In the sense of weak convergence,*

$$\lim_{r \rightarrow \infty} \mathcal{L}(\delta_r T^{(\delta_r)}) = \mathcal{M}_E(\lambda, a)$$

### Proof

In view of (12) and by definition of the laws  $\mathcal{M}_G(\theta^{(\delta)}, a^{(\delta)})$  and  $\mathcal{M}_E(\lambda, a)$ , it is sufficient to show that if for some fixed  $l > 0$ ,  $R^{(\delta)}$  is distributed according to  $\mathcal{G}(\exp(-\delta l))$ , with  $\delta > 0$ , then

$$\lim_{\delta \rightarrow 0_+} \mathcal{L}(\delta R^{(\delta)}) = \mathcal{L}(R)$$

where  $R$  is distributed according to an exponential distribution of parameter  $l$ . This convergence can be proven through the use of Laplace transforms, or from a more probabilistic point of view, by noticing that  $R^{(\delta)}$  has the same distribution as  $\lceil R/\delta \rceil$ , which denote the smallest integer larger or equal to  $R/\delta$ .

■

By resorting to Laplace transforms, one can see that the mapping

$$\mathcal{P}(\llbracket 0, N-1 \rrbracket) \ni a' \mapsto \mathcal{M}_E(\lambda, a') \in \mathcal{P}(\mathbb{R}_+)$$

is one to one, so we get a posteriori that

$$\lim_{\delta \rightarrow 0_+} a^{(\delta)} = a$$

To finish the proof of Theorem 2, we must show that  $a_0 > 0$ . Indeed, use similar arguments to those given in discrete time: it is well-known that for any initial distribution  $m_0$  on  $S$ ,

$$\lim_{t \rightarrow +\infty} \frac{1}{t} \ln(\mathbb{P}[T > t]) = -\lambda_0$$

(this can be easily deduced from the similar discrete time result (5) through finite skeletons). But if  $J = \min\{j \in \llbracket 0, N-1 \rrbracket : a_j > 0\}$ , and if  $T$  is distributed according to  $\mathcal{M}_{\mathcal{E}}(\lambda, a)$ , then

$$\forall 0 \leq l < \lambda_J, \quad \mathbb{E}[\exp(lT)] < +\infty$$

so from the fact that  $\lambda_0 < \lambda_1$ , we conclude that  $J = 0$ , i.e.  $a_0 > 0$ .

More structural results can be deduced by exhibiting an intertwining relation. Again start with skeleton subchains. For any  $\delta > 0$ , associate to the initial law  $m_0$  and to the absorbing transition matrix  $\bar{P}_\delta$ , a sequence  $(\mu_i^{(\delta)})_{i \in \llbracket 0, N-1 \rrbracket}$  of local equilibria on  $S$  and a sequence  $(d_i^{(\delta)})_{i \in \llbracket 0, N-2 \rrbracket}$  of real numbers from  $[0, 1]$ , with the convention that  $d_{N-1}^{(\delta)} = 1$ , as in Lemma 6. By compactness of  $\mathcal{P}(S)$  and of  $[0, 1]$ , we can find a sequence  $(\mu_i)_{i \in \llbracket 0, N-1 \rrbracket}$  of probabilities on  $S$ , a sequence  $(d_i)_{i \in \llbracket 0, N-1 \rrbracket}$  of real numbers from  $[0, 1]$  and a decreasing sequence  $(\delta^{(r)})_{r \in \mathbb{N}}$  of positive real numbers converging to 0, such that

$$\begin{aligned} \lim_{r \rightarrow \infty} (\mu_i^{(\delta^{(r)})})_{i \in \llbracket 0, N-1 \rrbracket} &= (\mu_i)_{i \in \llbracket 0, N-1 \rrbracket} \\ \lim_{r \rightarrow \infty} (d_i^{(\delta^{(r)})})_{i \in \llbracket 0, N-1 \rrbracket} &= (d_i)_{i \in \llbracket 0, N-1 \rrbracket} \end{aligned}$$

(in particular  $d_{N-1} = 1$ ).

Next introduce an absorbing generator  $\bar{K}$  on  $\bar{E} := \llbracket 0, N-1 \rrbracket \sqcup \{\triangle\}$  given by

$$\forall i, j \in \bar{E}, \quad \bar{K}(i, j) := \begin{cases} -\lambda_{N-1-i} & , \text{ if } i = j \in \llbracket 0, N-1 \rrbracket \\ d_i \lambda_{N-1-i} & , \text{ if } i \in \llbracket 0, N-1 \rrbracket \text{ and } j = \triangle \\ (1 - d_i) \lambda_{N-1-i} & , \text{ if } i \in \llbracket 0, N-2 \rrbracket \text{ and } j = i+1 \\ 1 & , \text{ if } i = j = \triangle \\ 0 & , \text{ otherwise} \end{cases} \quad (13)$$

and a Markov kernel  $\bar{\Lambda}$  from  $\bar{E}$  to  $\bar{S}$  given by

$$\forall i \in \bar{E}, \forall x \in \bar{S}, \quad \bar{\Lambda}(i, x) := \begin{cases} \mu_i(x) & , \text{ if } i \in \llbracket 0, N-1 \rrbracket \\ \delta_\triangle(x) & , \text{ if } i = \triangle \end{cases}$$

Their interest comes from the following intertwining relation

**Lemma 14** *We have the intertwining relation*

$$\Lambda \bar{L} = \bar{K} \Lambda$$

**Proof**

For  $\delta > 0$ , let  $\bar{Q}^{(\delta)}$  and  $\Lambda^{(\delta)}$  be the matrices associated to  $(\theta_i^{(\delta)})_{i \in \llbracket 0, N-1 \rrbracket}$ ,  $(\mu_i^{(\delta)})_{i \in \llbracket 0, N-1 \rrbracket}$  and  $(d_i^{(\delta)})_{i \in \llbracket 0, N-1 \rrbracket}$  as in the beginning of section 4. So

$$\Lambda^{(\delta)} \bar{P}_\delta = \bar{Q}^{(\delta)} \Lambda^{(\delta)}$$

and by consequence

$$\Lambda^{(\delta)} \frac{\bar{P}_\delta - \text{Id}}{\delta} = \frac{\bar{Q}^{(\delta)} - \text{Id}}{\delta} \Lambda^{(\delta)}$$

Now by replacing  $\delta$  by  $\delta_r$  and by letting  $r$  go to infinity, we get the needed relation. ■

Notice that the first row of  $\Lambda$  is  $m_0$ , since this is also true for  $\Lambda^{(\delta)}$ , for any  $\delta > 0$ . These facts lead us to consider a particular absorbing cadlag Markov process  $Y := (Y_t)_{t \geq 0}$  on  $\bar{E}$ , starting from 0 and whose generator is  $\bar{K}$ . Indeed, Fill [5] has shown that under the above intertwining relation and the equality  $\mathcal{L}(Y_0)\Lambda = \mathcal{L}(X_0)$ ,  $X$  and  $Y$  can be intertwined: there exists a coupling of  $X$  and  $Y$  such that the process  $(X_t, Y_t)_{t \geq 0}$  is Markovian, and such that for any time  $t \geq 0$ , the conditional law of  $X_t$  knowing  $(Y_s)_{0 \leq s \leq t}$  is given by

$$\mathcal{L}(X_t | (Y_s)_{0 \leq s \leq t}) = \Lambda(Y_t, \cdot)$$

As in discrete time, heuristically  $Y$  indicates which probability among  $\{\mu_i : i \in \llbracket 0, N-1 \rrbracket\}$  (whose elements can also be interpreted as local equilibria) has been reached by  $X$ , and in this respect, it is interesting to know that if a trajectory  $(X_t)_{t \geq 0}$  is given, it is possible to construct  $(Y_t)_{t \geq 0}$  from it in a progressive way, using also independent randomness (for more details, see Fill [10]).

Under the above coupling, let us denote

$$\begin{aligned} T_X &:= \inf\{t \geq 0 : X_t = \Delta\} \\ T_Y &:= \inf\{t \geq 0 : Y_t = \Delta\} \end{aligned}$$

It is clear that  $T_X$  has the same law as  $T$  and that  $T_Y$  is distributed according to  $\mathcal{M}_{\mathcal{E}}(\theta, a')$ , with the probability  $a'$  given by

$$\forall i \in \llbracket 0, N-1 \rrbracket, \quad a'_{N-1-i} := d_i \prod_{j \in \llbracket 0, i-1 \rrbracket} (1 - d_j)$$

The last assertion is a consequence of the traditional description of the evolution of a finite cadlag homogeneous Markov process using exponential times and the embedded Markov chain (which records the successive different positions taken by the process).

Next remark that the proof of Lemma 9 is also valid in continuous time, so under the above coupling  $T_X = T_Y$  a.s. In particular it follows that  $a' = a$  and since we have already seen that  $a_0 > 0$ , it follows that

$$\forall i \in \llbracket 0, N-2 \rrbracket, \quad d_i \in [0, 1)$$

We can deduce a posteriori that it was not necessary to take subsequences: first taking into account (11) (with  $I = 0$ ), we obtain

$$\lim_{\delta \rightarrow 0_+} (d_i^{(\delta)})_{i \in \llbracket 0, N-1 \rrbracket} = (d_i)_{i \in \llbracket 0, N-1 \rrbracket}$$

Next Lemma 14 implies that

$$\forall i \in \llbracket 0, N-1 \rrbracket, \quad \mu_i \bar{L} = -\lambda_{N-i-1} \mu_i + \lambda_{N-i-1} ((1 - d_i) \mu_{i+1} + d_i \delta_{\Delta}) \quad (14)$$

which shows that the local equilibria can be iteratively determined by this relation and the fact that  $\mu_0 = m_0$ . So it follows that

$$\lim_{\delta \rightarrow 0_+} (\mu_i^{(\delta)})_{i \in \llbracket 0, N-1 \rrbracket} = (\mu_i)_{i \in \llbracket 0, N-1 \rrbracket}$$

Finally, by similar arguments, it is also possible to extend Proposition 10. Let  $\bar{\mathcal{L}}$  be the closure of the set of generator matrices  $\bar{L}$  on  $\bar{S}$  whose unique recurrence class is  $\{\Delta\}$  and such that the restrictions  $L$  of  $\bar{L}$  to  $S \times S$  are irreducible and reversible. Next let  $\hat{\mathcal{L}}$  be the set of matrices  $\bar{L}$  in  $\bar{\mathcal{L}}$  such that  $\{\infty\}$  is the unique recurrence class of  $\bar{L}$ . If  $m_0 \in \mathcal{P}(S)$  and  $\bar{L} \in \hat{\mathcal{L}}$  are given, associate

to them a cadlag Markov process, starting from  $m_0$  with generator  $\bar{L}$ . Let us write  $\mathcal{L}(m_0, \bar{L})$  for the law on  $\mathbb{R}_+$  of the corresponding absorbtion time. Quite naturally, also introduce

$$\widehat{\mathcal{M}}_{\mathcal{E}} := \{\mathcal{M}_{\mathcal{E}}(\lambda, a) : \theta \in \widehat{\Lambda}, a \in \mathcal{P}(\llbracket 0, N-1 \rrbracket)\}$$

with

$$\widehat{\Lambda} := \{\lambda = (\lambda_i)_{i \in \llbracket 0, N-1 \rrbracket} \in (\mathbb{R}_+^*)^N : \lambda_0 \leq \lambda_1 \leq \lambda_2 \leq \dots \leq \lambda_{N-1}\}$$

Then we have

**Proposition 15** *The set of distributions  $\mathcal{L}(m_0, \bar{L})$ , when  $m_0$  runs through  $\mathcal{P}(S)$  and  $\bar{L}$  runs through  $\widehat{\mathcal{L}}$ , coincides with  $\widehat{\mathcal{M}}_{\mathcal{E}}$ .*

## 6 Negative eigenvalues

We return to the discrete time setting and investigate the peculiar features coming from the existence of negative eigenvalue(s). They will lead to an implicit description of the law of the absorbtion time involving simple probabilistic quantities and to an explicit formulation (but maybe less easy to manipulate) in terms of local equilibria and death proportions.

As in the introduction,  $\bar{P}$  is a Markov transition matrix on  $\bar{S}$  such that  $\{\Delta\}$  is the unique recurrence class and such that the restriction  $P$  of  $\bar{P}$  to  $S \times S$  is irreducible and reversible. So we can consider  $(\theta_i)_{i \in \llbracket 0, N-1 \rrbracket}$  the eigenvalues of  $P$  in nonincreasing order, a priori they belong to  $(-1, 1)$ . We are also given  $m_0$  an initial distribution on  $S$  and  $X$  a Markov chain starting from  $m_0$  whose transitions are described by  $\bar{P}$ . We are interested in the absorbtion time  $T_X := \inf\{n \in \mathbb{N} : X_n = \Delta\}$ . To describe its distribution, introduce the number of negative eigenvalues

$$I := \inf\{i \in \llbracket 0, N-1 \rrbracket : \theta_{N-1-i} \geq 0\}$$

and we assume that  $I \geq 1$  (by Perron-Frobenius theory, we know that  $\theta_0 \geq 0$ , so  $I \leq N-1$ ), namely that we are no longer in the situation where all the eigenvalues are nonnegative.

A first observation is that the beginning of the construction of the local equilibria is easier in this situation. Indeed, for any  $i \in \llbracket 0, I-1 \rrbracket$ ,  $(\bar{P} - \theta_{N-1-i} \text{Id}_{\bar{S}}) / (1 - \theta_{N-1-i})$  is a Markov transition matrix, as a convex combination of  $\bar{P}$  and the  $\bar{S} \times \bar{S}$  identity matrix  $\text{Id}_{\bar{S}}$ . So we don't need the result of Micchelli and Willoughby [21] to see that the family of measures  $(\tilde{m}_i)_{i \in \llbracket 0, I \rrbracket}$  defined by

$$\forall i \in \llbracket 0, I \rrbracket, \quad \tilde{m}_i := m_0 \prod_{j \in \llbracket 0, i-1 \rrbracket} \frac{\bar{P} - \theta_{N-1-j} \text{Id}_{\bar{S}}}{1 - \theta_{N-1-j}} \quad (15)$$

is in fact a family of probabilities on  $\bar{S}$ . As before, we can decompose them into

$$\forall i \in \llbracket 0, I \rrbracket, \quad \tilde{m}_i = \tilde{m}_i(S) \mu_i + \tilde{m}_i(\Delta) \delta_{\Delta}$$

where  $\mu_i$  is a probability on  $S$  (in particular,  $\mu_0 = m_0$ ). Since  $\Delta$  is absorbing, the map  $\llbracket 0, I \rrbracket \ni i \mapsto \tilde{m}_i(S)$  is nonincreasing, so we write

$$\forall i \in \llbracket 0, I \rrbracket, \quad \tilde{m}_i(S) = \prod_{j \in \llbracket 0, i-1 \rrbracket} (1 - d_j)$$

with  $(d_i)_{i \in \llbracket 0, I-1 \rrbracket}$  a family of numbers from  $[0, 1]$ . In fact it is not possible that for some  $i \in \llbracket 0, I-1 \rrbracket$ ,  $d_i = 1$ , otherwise we would have  $m_0 \bar{P}^i = \delta_{\Delta}$  and this would lead to a contradiction to the irreducibility of  $P$ . This implies that  $X$  can stay inside  $S$  for arbitrarily large times (but with an exponentially smaller and smaller probability as time is running on). The families  $(\mu_i)_{i \in \llbracket 0, I \rrbracket}$  of

probabilities on  $S$  and  $(d_i)_{i \in \llbracket 0, I \rrbracket}$  of numbers from  $[0, 1)$ , are uniquely determined by  $\mu_0 = m_0$  and the iterative relations

$$\forall i \in \llbracket 0, I \rrbracket, \quad \mu_i \bar{P} = \theta_{N-1-i} \mu_i + (1 - \theta_{N-1-i})((1 - d_i) \mu_{i+1} + d_i \delta_\Delta) \quad (16)$$

But since  $\theta_{N-1-i} < 0$  for  $i \in \llbracket 0, I-1 \rrbracket$ , Lemma 6 is of no help here. Another probabilistic interpretation is needed. The heuristic principle is that negative eigenvalues make the chain  $X$  run “too fast” and we have to slow it down. To do so, let  $(B_i)_{i \in \llbracket 0, I-1 \rrbracket}$  be independent Bernoulli random variables of respective parameters  $(1/(1 - \theta_{N-1-i}))_{i \in \llbracket 0, I-1 \rrbracket}$ :

$$\forall i \in \llbracket 0, I-1 \rrbracket, \quad \mathbb{P}[B_i = 1] = \frac{1}{1 - \theta_{N-1-i}} = 1 - \mathbb{P}[B_i = 0]$$

The family  $(B_i)_{i \in \llbracket 0, I-1 \rrbracket}$  is assumed to be independent of  $X$ . Next, define the time change

$$\forall i \in \mathbb{N}, \quad S(i) := \begin{cases} \sum_{j \in \llbracket 0, i-1 \rrbracket} B_j & , \text{ if } i \in \llbracket 0, I \rrbracket \\ S(I) + i - I & , \text{ if } i \geq I + 1 \end{cases}$$

which is independent of  $X$  and consider the chain  $Y := (Y_i)_{i \in \mathbb{N}}$  given by

$$\forall i \in \mathbb{N}, \quad Y_i := X_{S(i)}$$

It is in fact an inhomogeneous Markov chain with initial distribution  $m_0$  and transition matrix at time  $i \in \mathbb{N}$

$$\begin{cases} \frac{\bar{P} - \theta_{N-1-i} \text{Id}_{\bar{S}}}{1 - \theta_{N-1-i}} & , \text{ if } i \in \llbracket 0, I-1 \rrbracket \\ P & , \text{ if } i \geq I \end{cases}$$

Let  $T_Y := \inf\{n \in \mathbb{N} : Y_n = \Delta\}$  be the absorption time of  $Y$  (even if this chain is inhomogeneous). Its law is easy to describe, but we first need to extend the families  $(\mu_i)_{i \in \llbracket 0, I \rrbracket}$  and  $(d_i)_{i \in \llbracket 0, I \rrbracket}$  into  $(\mu_i)_{i \in \llbracket 0, N-1 \rrbracket}$  and  $(d_i)_{i \in \llbracket 0, N-1 \rrbracket}$  by using the iterative relations (16) equally for  $i \in \llbracket I+1, N-1 \rrbracket$ . For the same reason as above, for  $i \in \llbracket I+1, N-2 \rrbracket$ ,  $d_i \in [0, 1)$ , but to have that the  $\mu_i$  belong to  $\mathcal{P}(S)$ , for  $i \in \llbracket I+1, N-1 \rrbracket$ , we need to use the result of Micchelli and Willoughby [21] recalled in Theorem 5. Indeed, up to a normalizing factor,  $\mu_i$  is just  $m_0 \prod_{j \in \llbracket 0, i-1 \rrbracket} (P - \theta_{N-1-j} \text{Id}_S) / (1 - \theta_{N-1-j})$ . We also end up with  $d_{N-1} = 1$  and by convention we take  $\mu_N = \delta_\Delta$ .

Next consider  $\theta^+ := (\theta_i^+)_{i \in \llbracket 0, N-1 \rrbracket}$  (where for any  $r \in \mathbb{R}$ ,  $r^+ = r \vee 0$ ) and let  $a := (a_i)_{i \in \llbracket 0, N-1 \rrbracket}$  be the probability defined from the death proportions  $(d_i)_{i \in \llbracket 0, N-1 \rrbracket}$  as in (4), in particular  $a_0 > 0$ . The next assertion can be seen as a generalization of Theorem 1.

**Proposition 16** *The law of  $T_Y$  is  $\mathcal{M}_{\mathcal{G}}(\theta^+, a)$ .*

To prove this result, consider an intertwining between the inhomogeneous  $Y$  and a homogeneous Markov chain  $Z$  constructed as in the beginning of section 4, but with  $\theta$  replaced by  $\theta^+$ . More precisely,  $Z$  takes values in  $\bar{E} := \llbracket 0, N-1 \rrbracket \sqcup \{\Delta\}$ , starts from 0 and its transition matrix  $\tilde{Q}$  is given by

$$\forall i, j \in \bar{E}, \quad \tilde{Q}(i, j) := \begin{cases} \theta_{N-1-i}^+ & , \text{ if } i = j \in \llbracket 0, N-1 \rrbracket \\ d_i(1 - \theta_{N-1-i}^+) & , \text{ if } i \in \llbracket 0, N-1 \rrbracket \text{ and } j = \Delta \\ (1 - d_i)(1 - \theta_{N-1-i}^+) & , \text{ if } i \in \llbracket 0, N-2 \rrbracket \text{ and } j = i + 1 \\ 1 & , \text{ if } i = j = \Delta \\ 0 & , \text{ otherwise} \end{cases} \quad (17)$$

Again we denote by  $\Lambda$  the Markov kernel from  $\bar{E}$  to  $\bar{S}$  defined in (7). Then we have

**Lemma 17** *There exists a Markovian coupling between  $Y$  and  $Z$  such that for any time  $n \in \mathbb{N}$ , the conditional law of  $Y_n$  knowing  $Z_0, Z_1, \dots, Z_n$  is*

$$\mathcal{L}(Y_n | Z_0, Z_1, \dots, Z_n) = \Lambda(Z_n, \cdot)$$

**Proof**

We still have  $\mathcal{L}(Z_0)\Lambda = m_0 = \mathcal{L}(Y_0)$ , but since  $Y$  is not homogeneous, the intertwining cannot be reduced to an intertwining property between transition matrices.

First construct  $(Y_i, Z_i)_{i \in \llbracket 0, I \rrbracket}$  in the following way: Begin by considering  $(Y)_{i \in \llbracket 0, I \rrbracket}$  as before. Next, define

$$\forall i \in \llbracket 0, I \rrbracket, \quad Z_i := \begin{cases} i & , \text{ if } Y_i \in S \\ \Delta & , \text{ if } Y_i = \Delta \end{cases}$$

Let us check the assertion of the above lemma for  $n \in \llbracket 0, I \rrbracket$ . If  $Z_n = \Delta$ , then by construction and from the fact that  $\Delta$  is absorbing, we know that  $Y_n = \Delta$ , so the relation is satisfied. On the other hand, if  $Z_n \neq \Delta$ , then we just know that  $Y_n \neq \Delta$ . But (15) indicates that for  $n \in \llbracket 0, I \rrbracket$ , the law of  $Y_n$  is  $\tilde{m}_n$  and by construction  $\mu_n$  is the renormalization of the restriction of  $\tilde{m}_n$  to  $S$ , thus we get the relation of the lemma.

Next the construction of  $(Y_i, Z_i)_{i \in \llbracket I+1, +\infty \rrbracket}$  is more traditional. Indeed, let  $\hat{E} := \llbracket I, N-1 \rrbracket \sqcup \{\Delta\}$ ,  $\hat{Q}$  be the  $\hat{E} \times \hat{E}$ -restriction of  $\tilde{Q}$  and  $\hat{\Lambda}$  be the  $\hat{E} \times \bar{S}$ -restriction of  $\Lambda$ . These matrices are Markov kernels and we check from the extension of (16) to  $i \in \llbracket I, N-1 \rrbracket$  that we have

$$\hat{\Lambda} \bar{P} = \hat{Q} \hat{\Lambda}$$

(see the proof of Lemma 8). Since furthermore we have  $\mathcal{L}(Z_I)\hat{\Lambda} = \mathcal{L}(Y_I)$  with  $\mathcal{L}(Z_I) = \tilde{m}_I(S)\delta_I + \tilde{m}_I(\Delta)\delta_\Delta \in \mathcal{P}(\hat{E})$ , the arguments of Diaconis and Fill [5] enables us to construct a Markov chain  $(\hat{Y}_i, \hat{Z}_i)_{i \in \llbracket I+1, +\infty \rrbracket}$  such that

- we have  $(\hat{Y}_I, \hat{Z}_I) = (Y_I, Z_I)$
- $(\hat{Y}_i)_{i \in \llbracket I+1, +\infty \rrbracket}$  is a homogeneous Markov chain with transitions described by  $\bar{P}$
- $(\hat{Z}_i)_{i \in \llbracket I+1, +\infty \rrbracket}$  is a homogeneous Markov chain with transitions described by  $\hat{Q}$ , or equivalently by  $\tilde{Q}$
- at any time  $n \in \llbracket I, +\infty \rrbracket$ , the conditional law of  $\hat{Y}_n$  knowing  $\hat{Z}_I, \hat{Z}_{I+1}, \dots, \hat{Z}_n$  is

$$\begin{aligned} \mathcal{L}(\hat{Y}_n | \hat{Z}_I, \hat{Z}_{I+1}, \dots, \hat{Z}_n) &= \hat{\Lambda}(\hat{Z}_n, \cdot) \\ &= \Lambda(\hat{Z}_n, \cdot) \end{aligned}$$

These facts lead us to define  $(Y_i, Z_i)_{i \in \llbracket I+1, +\infty \rrbracket} := (\hat{Y}_i, \hat{Z}_i)_{i \in \llbracket I+1, +\infty \rrbracket}$ , because we then easily deduce the properties announced in the above lemma. In particular, notice that for  $n \in \llbracket I, +\infty \rrbracket$ , to condition  $Y_n$  with respect to  $(Z_0, Z_1, \dots, Z_n)$  is the same as to condition with respect to  $(Z_I, Z_{I+1}, \dots, Z_n)$  (since  $(Z_0, Z_1, \dots, Z_I)$  is deterministic if  $Z_n \neq \Delta$ ). ■

Now it is easy to deduce Proposition 16. Indeed, the proof of Lemma 9 is still valid and shows that  $\mathcal{L}(T_Y) = \mathcal{L}(T_Z)$ . It is quite clear that  $T_Z$  is distributed as  $\mathcal{M}_{\mathcal{G}}(\theta^+, a)$ . We remark that the families  $(\mu_i)_{i \in \llbracket 0, N-1 \rrbracket}$  and  $(d_i)_{i \in \llbracket 0, N-1 \rrbracket}$  retain their interpretations as local equilibria and death proportions, but with respect to the speeded up Markov chain  $Y$ , not  $X$ .

But we are interested in  $T_X$ , not in  $T_Y$ . By construction of  $Y$ , it appears that  $T_X = S(T_Y)$ , nevertheless this relation does not seem easy to exploit in general, since  $S$  and  $T_Y$  are not independent. Let  $S^{-1}$  be the generalized inverse of  $S$ :

$$\forall n \in \mathbb{N}, \quad S^{-1}(n) := \inf \{k \in \mathbb{N} : S(k) \geq n\}$$



Then it is immediate to see that we also have

$$T_Y = S^{-1}(T_X) \quad (18)$$

and since  $T_X$  and  $S^{-1}$  are independent, this equality and Proposition 16 implicitly determine the law of  $T_X$ :

**Lemma 18** *The law  $\mathcal{L}(T_X) \in \mathcal{P}(\mathbb{N}^*)$  is uniquely determined by the fact that if  $T$  is distributed as  $\mathcal{L}(T_X)$  and is independent of  $S$ , then  $S^{-1}(T)$  has  $\mathcal{M}_G(\theta^+, a)$  as law.*

**Proof**

It is sufficient to write that for any  $n \in \mathbb{N}^*$ , we have

$$\begin{aligned} \mathcal{M}_G(\theta^+, a)[n] &= \mathbb{P}[S^{-1}(T) = n] \\ &= \sum_{k \in \llbracket 1, n \rrbracket} \mathbb{P}[S^{-1}(k) = n, T = k] \\ &= \sum_{k \in \llbracket 1, n \rrbracket} \mathbb{P}[S^{-1}(k) = n] \mathbb{P}[T = k] \end{aligned}$$

(notice that for any  $k \in \mathbb{N}$ ,  $S^{-1}(k) \geq k$ ), and since for any  $n \in \mathbb{N}^*$ ,  $\mathbb{P}[S^{-1}(n) = n] > 0$  these identities enable to compute iteratively  $\mathbb{P}[T = n]$  for any  $n \in \mathbb{N}^*$ . ■

We could deduce from the above proof an explicit formula for the law of  $T_X$ , but there is a simpler way to deduce it.

**Proposition 19** *For any  $n \in \mathbb{N}^*$ ,*

$$\mathbb{P}[T_X = n] = \sum_{k \in \llbracket 1, n \rrbracket} b_{n,k} \mathcal{M}_G(\theta^+, a)[k] \quad (19)$$

where for any  $k \in \llbracket 1, n \rrbracket$ ,

$$b_{n,k} := \prod_{j \in \llbracket 0, (k-1) \wedge (I-1) \rrbracket} (1 - \theta_{N-1-j}) \sum_{i_0 + \dots + i_{(k-1) \wedge (I-1)} = n-k} \theta_{N-1}^{i_0} \theta_{N-2}^{i_1} \dots \theta_{N-1-(k-1) \wedge (I-1)}^{i_{(k-1) \wedge (I-1)}}$$

In particular  $(b_{n,k})_{n \in \mathbb{N}^*, k \in \llbracket 1, n \rrbracket}$  only depends on the negative eigenvalues, while  $\mathcal{M}_G(\theta^+, a)$  is defined in terms of the nonnegative eigenvalues.

**Proof**

To simplify the notation, let us denote for any  $k \in \mathbb{N}$ ,  $\hat{\theta}_k = \theta_{N-1-k} \wedge 0$  (with the convention that  $\hat{\theta}_k = 0$  for all  $k \geq N$ ). Next consider the polynomial in the  $z$  variable defined by

$$Q_k(z) := \prod_{j \in \llbracket 0, k-1 \rrbracket} \frac{z - \hat{\theta}_j}{1 - \hat{\theta}_j}$$

Then write for any  $n \in \mathbb{N}$ ,

$$z^n = \sum_{k \in \llbracket 0, n \rrbracket} c_{n,k} Q_k(z) \quad (20)$$

with for any  $k \in \llbracket 0, n \rrbracket$ ,

$$c_{n,k} := \prod_{j \in \llbracket 0, k-1 \rrbracket} (1 - \hat{\theta}_j) \sum_{i_0 + \dots + i_k = n-k} \hat{\theta}_0^{i_0} \hat{\theta}_1^{i_1} \dots \hat{\theta}_k^{i_k} \quad (21)$$

(in particular  $c_{0,0} = 1$ ). To get these formulas, use the observations that for any  $n \in \mathbb{N}$  and  $k \in \llbracket 0, n \rrbracket$ ,  $z^{n+1} = zz^n$  and  $XQ_k(z) = (1 - \hat{\theta}_k)Q_{k+1}(z) + \hat{\theta}_k Q_k(z)$ , which can be translated into the iterative relations

$$\begin{aligned} c_{n+1,0} &= c_{n,0} \hat{\theta}_0 \\ c_{n+1,k} &= c_{n,k-1}(1 - \hat{\theta}_{k-1}) + c_{n,k} \hat{\theta}_k, \quad \text{for } k \in \llbracket 1, n \rrbracket \\ c_{n+1,n+1} &= c_{n,n}(1 - \hat{\theta}_n) \end{aligned}$$

If the  $\theta_n$ , for  $n \in \mathbb{N}$ , were nonnegative, we would recognize that  $c_{n,k} = \mathbb{P}[R_n = k]$ , where  $(R_n)_{n \in \mathbb{N}}$  is a Markov chain on  $\mathbb{N}$  starting from 0 whose transitions are described by

$$\forall n \in \mathbb{N}, \quad \mathbb{P}[R_{n+1} = R_n + 1 | R_n] = 1 - \hat{\theta}_{R_n} = 1 - \mathbb{P}[R_{n+1} = R_n | R_n]$$

and (21) would follow without difficulty. But conversely, one can directly check by an elementary combinatorial argument that the above iterative relations are satisfied if, for  $n \in \mathbb{N}$  and  $k \in \llbracket 0, n \rrbracket$ , we take  $c_{n,k}$  as defined in (21), without any assumption on the signs of the  $\theta_n$ , for  $n \in \mathbb{N}$ .

Next we can translate (20) into information about absorbtion time: for any time  $n \in \mathbb{N}^*$ , we have

$$\begin{aligned} \mathbb{P}[T_X \leq n] &= \mathbb{P}[X_n = \triangle] \\ &= m_0 \bar{P}^n \mathbb{1}_{\{\triangle\}} \\ &= m_0 \left( \sum_{k \in \llbracket 1, n \rrbracket} c_{n,k} Q_k(\bar{P}) \right) \mathbb{1}_{\{\triangle\}} \\ &= \sum_{k \in \llbracket 1, n \rrbracket} c_{n,k} m_0 Q_k(\bar{P}) \mathbb{1}_{\{\triangle\}} \\ &= \sum_{k \in \llbracket 1, n \rrbracket} c_{n,k} \mathbb{P}[Y_k = \triangle] \\ &= \sum_{k \in \llbracket 1, n \rrbracket} c_{n,k} \mathbb{P}[T_Y \leq k] \end{aligned}$$

So using that for any  $n \in \mathbb{N}^*$ ,  $\mathbb{P}[T = n] = \mathbb{P}[T \leq n] - \mathbb{P}[T \leq n - 1]$ , with  $T = T_X$  or  $T = T_Y$ , we obtain

$$\mathbb{P}[T_X = n] = \sum_{k \in \llbracket 1, n \rrbracket} b_{n,k} \mathbb{P}[T_Y = k]$$

with for any  $k \in \llbracket 1, n \rrbracket$ ,

$$b_{n,k} := c_{n,n} + \sum_{j \in \llbracket k, n-1 \rrbracket} c_{n,j} - c_{n-1,j}$$

To recover the formula given in the proposition, we notice that by the above iterative relations,

$$c_{n,j} - c_{n-1,j} = c_{n-1,j-1}(1 - \hat{\theta}_{j-1}) - c_{n-1,j}(1 - \hat{\theta}_j)$$

so

$$\begin{aligned} b_{n,k} &= c_{n,n} - c_{n-1,n-1}(1 - \hat{\theta}_{n-1}) + c_{n-1,k-1}(1 - \hat{\theta}_{k-1}) \\ &= c_{n-1,k-1}(1 - \hat{\theta}_{k-1}) \end{aligned}$$

We get the announced result, by taking into account that  $\hat{\theta}_j = 0$  for  $j \geq I$ . ■

One can give a reduced formula for the  $(b_{n,k})_{n \in \mathbb{N}^*, k \in \llbracket 1, n \rrbracket}$  appearing in the above proposition, at least if all the negative eigenvalues are distinct. For  $n \in \mathbb{N}$ , let  $p_n(x) := x^n$  be the  $n^{\text{th}}$  monomial, then in terms of divided difference, we have for any  $k \in \llbracket 0, n \rrbracket$ , and any real numbers  $x_0 < x_1 < \dots < x_k$ ,

$$\begin{aligned} \sum_{i_0 + \dots + i_k = n} x_0^{i_0} x_1^{i_1} \dots x_k^{i_k} &= p_{n+k}[x_0, x_1, \dots, x_k] \\ &= \sum_{i \in \llbracket 0, k \rrbracket} \frac{x_i^{n+k}}{\prod_{j \in \llbracket 0, k \rrbracket \setminus \{i\}} x_i - x_j} \end{aligned} \quad (22)$$

thus for any  $n \in \mathbb{N}^*$  and  $k \in \llbracket 1, n \rrbracket$ ,

$$b_{n,k} = \prod_{j \in \llbracket 0, (k-1) \wedge (I-1) \rrbracket} (1 - \theta_{N-1-j}) \sum_{i \in \llbracket 0, (k \wedge I) - 1 \rrbracket} \frac{\theta_{N-1-i}^{n-k-1+k \wedge I}}{\prod_{j \in \llbracket 0, (k \wedge I) - 1 \rrbracket \setminus \{i\}} \theta_{N-1-i} - \theta_{N-1-j}}$$

One has to be careful about the decomposition of the law of  $T_X$  given in Proposition 19, since some of the terms of the sum can be negative. Indeed,  $b_{n,k}$  is nonpositive if  $n - k$  is odd and nonnegative otherwise. Nevertheless, it seems that some partial sums are positive, see the third point in the remarks below.

### Remarks 20

1) The probability  $a = (a_i)_{i \in \llbracket 0, N-1 \rrbracket}$  which appears in Propositions 16 and 19 satisfies  $a_i > 0 \Rightarrow a_{i-1} > 0$ , for any  $i \in \llbracket N - I, N - 1 \rrbracket$ . In fact, if  $a_i > 0$  then  $d_{N-1-i} > 0$ , which means there exists an element  $x \in S$  in the support of  $\tilde{m}_{N-1-i}$  such that  $\bar{P}(x, \Delta) > 0$ , but since  $\tilde{m}_{N-i}$  is the transportation of  $\tilde{m}_{N-1-i}$  through the Markov matrix  $(\bar{P} - \theta_i \text{Id}_{\bar{S}})/(1 - \theta_i)$  whose diagonal is positive,  $x$  also belongs to the support of  $\tilde{m}_{N-i}$ , so  $d_{N-i} > 0$  and  $a_{i-1} > 0$ . We believe that this condition may be true in general for  $i \in \llbracket 1, N - 1 \rrbracket$ , that is why we alluded to it in the previous section. It would be a consequence of the property that the mapping associating to  $i \in \llbracket 0, N - 1 \rrbracket$  the support of  $\mu_i$  is nondecreasing.

2) Let  $\Theta_{\text{SM}}$  be the set of eigenvalues  $(\theta_i)_{i \in \llbracket 0, N-1 \rrbracket}$  in nonincreasing order of irreducible and reversible  $N \times N$  subMarkovian matrices. It seems it is not easy to characterize  $\Theta_{\text{SM}}$ , since this is related to the famous nonnegative inverse eigenvalue problem, see for instance the overview given by Egleston, Lenker, Terry and Narayan [9].

3) Suppose given a family  $\theta = (\theta_i)_{i \in \llbracket 0, N-1 \rrbracket} \in \Theta_{\text{SM}}$  and a probability  $a = (a_i)_{i \in \llbracket 0, N-1 \rrbracket}$  on  $\llbracket 0, N - 1 \rrbracket$ , satisfying that for any  $i \in \llbracket 1, N - 1 \rrbracket$ ,  $a_i > 0 \Rightarrow a_{i-1} > 0$  (or equivalently if  $a$  is given by (4), that for any  $i \in \llbracket 0, N - 2 \rrbracket$ ,  $d_i > 0 \Rightarrow d_{i+1} > 0$ ). In a similar spirit to what we have done in the previous section and at least under the above restriction on  $a$ , it seems possible that we can find a Markov chain as in the present section, whose absorption time is distributed as in Proposition 16. We assume for the remainder of these remarks that such a Markov chain exists, denoted  $X^{(\theta, a)}$ . For simplicity, we also make the assumption that the death proportions  $(d_i)_{i \in \llbracket 0, N-1 \rrbracket}$  associated to our original Markov chain  $X$  satisfy the condition  $\forall i \in \llbracket 0, N - 2 \rrbracket$ ,  $d_i > 0 \Rightarrow d_{i+1} > 0$ .

4) Let  $J \geq I$  be defined by

$$J := \inf\{i \in \llbracket 0, N - 1 \rrbracket : \theta_{N-1-i} > 0\}$$

and for any  $j \in \llbracket 0, J \rrbracket$ , let us decompose the sum of Proposition 16 into

$$\mathbb{P}[T_X = n] = \sum_{k \in \llbracket 1, n \wedge j \rrbracket} b_{n,k} \mathcal{M}_{\mathcal{G}}(\theta^+, a)[k] + \sum_{k \in \llbracket j+1, n \rrbracket} b_{n,k} \mathcal{M}_{\mathcal{G}}(\theta^+, a)[k] \quad (23)$$

Then the last sum is nonnegative and it can be written in the following form

$$\sum_{k \in \llbracket j+1, n \rrbracket} b_{n,k} \mathcal{M}_{\mathcal{G}}(\theta^+, a)[k] = \mathcal{M}_{\mathcal{G}}(\theta^+, a)[\llbracket j+1, \infty \rrbracket] \mathbb{P}[T_{X^{(\theta, \hat{a})}} = n] \quad (24)$$

where the probability  $\check{a}$  is defined as  $a$ , but in terms of the modified death proportions  $\check{d} := (\check{d}_i)_{i \in \llbracket 0, N-1 \rrbracket}$  given by

$$\forall i \in \llbracket 0, N-1 \rrbracket, \quad \check{d}_i := \begin{cases} 0 & , \text{ if } i \in \llbracket 0, j-1 \rrbracket \\ d_i & , \text{ otherwise} \end{cases}$$

Let us also denote  $d = (d_i)_{i \in \llbracket 0, N-1 \rrbracket}$  and write  $Z^{(d)}$  for the chain  $Z$  constructed after Proposition 16, to indicate explicitly its dependence on the vector  $d$ . From the fact that  $j \in \llbracket 0, J \rrbracket$ , it follows that

$$\forall k \in \llbracket j+1, \infty \rrbracket, \quad \mathbb{P}[T_{Z^{(d)}} = k | T_{Z^{(d)}} \geq j+1] = \mathbb{P}[T_{Z^{(\check{d})}} = k]$$

and this can be rewritten in the form

$$\begin{aligned} \forall k \in \llbracket j+1, \infty \rrbracket, \quad \mathbb{P}[T_{Z^{(d)}} = k] &= \mathbb{P}[T_{Z^{(d)}} \geq j+1] \mathbb{P}[T_{Z^{(\check{d})}} = k] \\ &= \mathcal{M}_{\mathcal{G}}(\theta^+, a)[\llbracket j+1, \infty \rrbracket] \mathcal{M}_{\mathcal{G}}(\theta^+, \check{a})[k] \end{aligned}$$

Furthermore we have for any  $k \in \llbracket 0, j \rrbracket$ ,  $\mathcal{M}_{\mathcal{G}}(\theta^+, \check{a})[k] = 0$ , so if we apply Proposition 16 to the Markov chain  $X^{(\theta, \check{a})}$ , we get (24).

5) In the above considerations, let us take  $j = I$ . It is tempting to believe that the conditional law of  $T_X$  knowing that  $T_Y > I$  is the law of  $T_{X^{(\theta, \check{a})}}$ , so we can identify the quantity in (24) as  $\mathbb{P}[T_X = n, T_Y > I]$  and it would follow that the first sum in the r.h.s. of (23) (with  $j = I$ ) is equal to  $\mathbb{P}[T_X = n, T_Y \leq I]$ . But this is not true, since the latter sum can be negative. Indeed, assume that  $d_I > 0$  and let  $\tilde{a}$  be the probability on  $\llbracket 0, N-1 \rrbracket$  defined as  $a$ , but in terms of the modified death proportions  $\tilde{d} := (\tilde{d}_i)_{i \in \llbracket 0, N-1 \rrbracket}$  given by

$$\forall i \in \llbracket 0, N-1 \rrbracket, \quad \tilde{d}_i := \begin{cases} 0 & , \text{ if } i \in \llbracket 0, I-2 \rrbracket \\ 1/2 & , \text{ if } i = I-1 \\ d_i & , \text{ otherwise} \end{cases}$$

Then we get that  $\mathcal{M}_{\mathcal{G}}(\theta^+, \tilde{a})[k] = 0$  for any  $k \in \llbracket 1, I-1 \rrbracket$  and  $\mathcal{M}_{\mathcal{G}}(\theta^+, \tilde{a})[I] > 0$ , so that for  $n \geq I$ , the sign of

$$\sum_{k \in \llbracket 1, n \wedge I \rrbracket} b_{n,k} \mathcal{M}_{\mathcal{G}}(\theta^+, a)[k] = b_{n,I} \mathcal{M}_{\mathcal{G}}(\theta^+, a)[I]$$

is  $(-1)^{n-I}$ .

□

There is one situation where the implicit formula (18) can be simplified. Put an unoriented graph structure on  $\bar{S}$  by saying that  $x, y \in \bar{S}$  are neighbours if  $\bar{P}(x, y) > 0$  or  $\bar{P}(y, x) > 0$ . We denote by  $\rho$  the corresponding distance and assume that the support of  $m_0$  is included in  $\{x \in S : \rho(x, \Delta) > I\}$ . Then we have that  $d_i = 0$  for all  $i \in \llbracket 0, I-1 \rrbracket$  (this is equivalent to the above condition), so a.s.  $T_Y > I$  and we can write

$$T_Y = T_X + I - S(I) \tag{25}$$

Let  $\psi$  be the moment generating function of the law  $\mathcal{M}_{\mathcal{G}}(\theta^+, a)$  and  $\varphi$  be the moment generating function of  $I - S(I)$  (which is a sum of  $N$  independent Bernoulli random variables of respective parameters  $(-\theta_{N-1-i}/(1-\theta_{N-1-i}))_{i \in \llbracket 0, I-1 \rrbracket}$ ). Then the moment generating function of  $T_X$  is given by

$$\forall u \in [0, 1], \quad \mathbb{E}[u^{T_X}] = \frac{\psi(u)}{\varphi(u)} \tag{26}$$

We will encounter such a situation in next section.

More generally, it is natural to try to decompose  $T_X$  as a mixture by conditioning with respect to the  $\sigma$ -field generated by  $\{T_X > S(I)\} = \{T_Y > I\}$ . Indeed, conditioned by the event  $\{T_Y > I\}$ ,  $(Y_{I+n})_{n \in \mathbb{N}}$  is a homogeneous Markov chain with initial distribution  $\mu_I$  and transition kernel is  $\bar{P}$ . Thus according to the proof of Lemma 17, the law of  $T_Y - I$  conditioned by  $\{T_Y > I\}$  is the distribution of the absorption time for a Markov chain  $Z$  on  $\llbracket I, N-1 \rrbracket \sqcup \{\Delta\}$  starting from  $I$  and whose transitions are described by the restriction to  $(\llbracket I, N-1 \rrbracket \sqcup \{\Delta\})^2$  of the matrix  $\tilde{Q}$  defined in (17). So we have

$$\mathcal{L}(T_X - S(I) | T_X > S(I)) = \mathcal{M}_{\mathcal{G}}(\check{\theta}, \check{a})$$

where  $\check{\theta} = (\theta_i)_{i \in \llbracket 0, N-1-I \rrbracket}$  and  $\check{a}$  is the probability on  $\llbracket 0, N-1-I \rrbracket$  defined by

$$\forall i \in \llbracket 0, N-1-I \rrbracket, \quad \check{a}_{N-1-I-i} := d_{I+i} \prod_{j \in \llbracket I, I+i-1 \rrbracket} (1 - d_j)$$

Unfortunately the conditional law  $\mathcal{L}(T_X | T_X > S(I))$  is not easy to deduce, because on  $\{T_X > S(I)\}$ ,  $T_X$  and  $S(I)$  are no longer independent in general. Let us introduce the Markov kernel  $K$  from  $\mathbb{N}^*$  to  $\llbracket 0, I \rrbracket$  describing the conditional law  $\mathcal{L}(S(I) | T_X - S(I))$ . Let us also consider  $\nu$  the probability on  $\llbracket 0, I \rrbracket$  which is the distribution of  $T_X$  conditioned on  $\{T_X \leq S(I)\}$ . Then to sample according to  $\mathcal{L}(T_X)$  one has to do the following: first to draw a Bernoulli variable  $B$  of parameter  $\tilde{m}_I(S) = \prod_{j \in \llbracket 0, I-1 \rrbracket} (1 - d_j)$ . If  $B = 1$ , one samples  $T$  according to  $\mathcal{M}_{\mathcal{G}}(\check{\theta}, \check{a})$  and adds to  $T$  a variable distributed according to  $K(T, \cdot)$ . If  $B = 0$ , one samples according to  $\nu$ . In particular, it appears that  $T_X$  is stochastically dominated by  $I + \tilde{T}$ , where  $\tilde{T}$  is distributed according to the mixture

$$\left(1 - \prod_{j \in \llbracket 0, I-1 \rrbracket} (1 - d_j)\right) \delta_0 + \prod_{j \in \llbracket 0, I-1 \rrbracket} (1 - d_j) \mathcal{M}_{\mathcal{G}}(\check{\theta}, \check{a})$$

But a better bound can be obtained by noticing directly that  $T_X$  is stochastically dominated by  $T_Y$ . Both of these bounds are independent of the negative eigenvalues.

## 7 Examples

We present three illustrations of the previous results: finite birth and death chains which are absorbed at one end of their state space segment, the situation of constant probability/rate of absorption and the case  $N = 2$  in continuous time. Their common point is that the death proportions can be easily computed (at least when the initial point is the non-absorbing end in the birth and death setting). We hope to consider in future studies more serious applications, where estimates of death proportions are more involved (even if from a complexity point of view, they only ask for the knowledge of  $(m_0 \bar{P}^n(\Delta))_{n \in \llbracket 1, N-1 \rrbracket}$ , once the Dirichlet eigenvalues are known).

- We begin with the birth and death setting, which was the first motivation for this investigation (see [6] and in particular the end of its introduction for historical background). We take  $S = \llbracket 0, N-1 \rrbracket$  and write  $\Delta = N$ , where  $N \in \mathbb{N}^*$ .

First consider the discrete time situation, where the absorbing transition matrix  $\bar{P}$  satisfies

$$\forall i, j \in \llbracket 0, N \rrbracket, \quad \bar{P}(i, j) > 0 \iff |i - j| = 1 \text{ and } i \neq N$$

Then its restriction  $P$  to  $\llbracket 0, N-1 \rrbracket^2$  is irreducible and reversible, with respect to the measure  $\pi$  defined by

$$\forall i \in \llbracket 0, N-1 \rrbracket, \quad \pi(i) := \prod_{j \in \llbracket 0, i-1 \rrbracket} \frac{P(j, j+1)}{P(j+1, j)} \quad (27)$$

As usual, let  $(\theta_i)_{i \in \llbracket 0, N-1 \rrbracket}$  be the eigenvalues of  $P$ , in decreasing order (here they are all of multiplicity 1). Then we recover the following well-known result:

**Corollary 21** *Assume that the eigenvalues are nonnegative, i.e.  $\theta_{N-1} \geq 0$ , and that the initial distribution  $m_0$  is  $\delta_0$ . Then  $T_X$  is distributed as a sum of independent geometric distributions of respective parameters the  $\theta_i$ , for  $i \in \llbracket 0, N-1 \rrbracket$ .*

Indeed, for any  $n \in \llbracket 0, N-1 \rrbracket$ , the support of  $\bar{P}^n(0, \cdot)$  is included in  $\llbracket 0, n \rrbracket$ , so Lemma 6 shows that the death proportions  $d_i$ , for  $i \in \llbracket 0, N-2 \rrbracket$ , are null. In particular, we get that the probability  $a$  defined by (4) is just  $\delta_0$ . So Proposition 21 is an immediate consequence of Theorem 1.

The above arguments can also be applied when there are some negative eigenvalues (see equation (16)). Taking into account the discussion after Remarks 20, in particular (25) and (26), we get

**Corollary 22** *Without any assumptions on the sign of the eigenvalues, the moment generating function of  $T_X$ , when  $X$  is starting from 0, is given by*

$$\forall u \in [0, 1], \quad \mathbb{E}[u^{T_X}] = \prod_{i \in \llbracket 0, N-1 \rrbracket} \frac{(1 - \theta_i)u}{1 - \theta_i u}$$

*This has the probabilistic interpretation that if we add to  $T_X$  an independent sum of independent Bernoulli random variables of respective parameters  $(\theta_i^- / (1 + \theta_i^-))_{i \in \llbracket 0, N-1 \rrbracket}$  then we get a random variable distributed as a sum of independent geometric random variables of parameters  $(\theta_i^+)_{i \in \llbracket 0, N-1 \rrbracket}$ .*

Let us come back to the case of nonnegative eigenvalues. For more general initial distributions  $m_0 \in \mathcal{P}(\llbracket 0, N-1 \rrbracket)$  than  $\delta_0$ , the law of  $T_X$  is a “true” mixture of sum of geometric distributions (which can sometimes be simplified, for instance if  $m_0$  is the quasi-stationary law, we get a geometric law of parameter  $\theta_0$ , see Remark 7). Indeed, if  $I_0 := \max\{i \in \llbracket 0, N-1 \rrbracket : m_0(i) > 0\}$ , then  $\mathcal{L}(T_X)$  is a mixture of the distributions  $\mathcal{G}(\theta_0, \theta_1, \dots, \theta_{N-1})$ ,  $\mathcal{G}(\theta_1, \theta_2, \dots, \theta_{N-1})$ , ...,  $\mathcal{G}(\theta_{I_0}, \theta_{I_0+1}, \dots, \theta_{N-1})$ , with the notation of the introduction, since we must have  $d_i = 0$  for  $i \in \llbracket 0, N-2-I_0 \rrbracket$ . The situation becomes more involved if there exists some negative eigenvalues. But if we have  $I_0 < N-1-I$ , where  $I$  is the number of negative eigenvalues, then according to (25), there exists a probability  $a = (a_i)_{i \in \llbracket 0, I_0 \rrbracket}$  such that

$$\forall u \in [0, 1], \quad \mathbb{E}[u^{T_X}] = \sum_{i \in \llbracket 0, I_0 \rrbracket} a_i \prod_{j \in \llbracket i, N-1 \rrbracket} \frac{(1 - \theta_j)u}{1 - \theta_j u}$$

Let us now deal with the continuous time situation, which is simpler. Suppose  $\bar{L}$  is a generator on  $\llbracket 0, N \rrbracket$  satisfying

$$\forall i, j \in \llbracket 0, N \rrbracket, \quad \bar{L}(i, j) > 0 \iff |i - j| = 1 \text{ and } i \neq N$$

As above, its restriction  $L$  to  $\llbracket 0, N-1 \rrbracket^2$  is irreducible and reversible with respect to the measure  $\pi$  defined as in (27), with  $P$  replaced by  $L$ . We denote by  $(\lambda_i)_{i \in \llbracket 0, N-1 \rrbracket}$  the eigenvalues of  $-L$  in increasing order. Then we have the equivalent of Proposition 21:

**Corollary 23** *Assume that the Markov process  $X$  starts from 0. Then  $T_X$  is distributed as a sum of independent exponential distributions of respective parameters the  $\lambda_i$ , for  $i \in \llbracket 0, N-1 \rrbracket$ .*

This is an immediate consequence of the recursive formula (14), which implies that the death proportions  $d_i$  are null for  $i \in \llbracket 0, N-2 \rrbracket$ . We also recover the local equilibria that were considered in [6]:

$$\forall i \in \llbracket 0, N-1 \rrbracket, \quad \mu_i = \delta_0 \prod_{j \in \llbracket 0, i-1 \rrbracket} \frac{L - \lambda_{N-1-i} \text{Id}_{\llbracket 0, N-1 \rrbracket}}{\lambda_{N-1-i}}$$

even if they were not described so explicitly there. As above, for more general initial distribution  $m_0 \in \mathcal{P}(\llbracket 0, N-1 \rrbracket)$ ,  $\mathcal{L}(T_X)$  is a mixture of the sums of exponential distributions  $\mathcal{E}(\lambda_0, \lambda_1, \dots, \lambda_{N-1})$ ,  $\mathcal{E}(\lambda_1, \lambda_2, \dots, \lambda_{N-1})$ , ...,  $\mathcal{E}(\lambda_{I_0}, \lambda_2, \dots, \lambda_{N-1})$ , where  $I_0 := \max\{i \in \llbracket 0, N-1 \rrbracket : m_0(i) > 0\}$ .

• We now come to the example of constant absorbtion probability. It is characterized by the existence of  $a \in (0, 1)$  such that

$$\forall x \in S, \quad \bar{P}(x, \Delta) = 1 - a$$

This is equivalent to the fact that  $P$  can be written under the form  $P = a\tilde{P}$ , where  $\tilde{P}$  is a Markov transition matrix on  $S$ . Under our usual assumption that  $P$  is irreducible and reversible,  $\tilde{P}$  is also irreducible and reversible. Let  $(\tilde{\theta}_i)_{i \in \llbracket 0, N-1 \rrbracket}$  be its eigenvalues in nonincreasing order, which now belong to  $[-1, 1]$ . Thus

$$\forall i \in \llbracket 0, N-1 \rrbracket, \quad \theta_i = a\tilde{\theta}_i$$

and since  $\tilde{\theta}_0 = 1$ , we get  $a = \theta_0$ .

In this situation, whatever the initial distribution  $m_0 \in \mathcal{P}(S)$ , the absorbtion time  $T_X$  is distributed as the geometric law  $\mathcal{G}(\theta_0)$ . This is an immediate consequence of the fact that at each time  $n \in \mathbb{N}$ , if the chain has not yet been absorbed, then it has a chance  $1 - \theta_0$  to be absorbed at next step, independent of the underlying position. Taking into account (16) (for all  $i \in \llbracket 0, N-1 \rrbracket$ ), it follows that we have for the death proportions

$$\forall i \in \llbracket 0, N-1 \rrbracket, \quad d_i = \frac{1 - \theta_0}{1 - \theta_{N-1-i}}$$

Let  $a = (a_i)_{i \in \llbracket 0, N-1 \rrbracket}$  be the weight given by (4):

$$\forall i \in \llbracket 0, N-2 \rrbracket, \quad a_i = \frac{1 - \theta_0}{1 - \theta_i} \prod_{j \in \llbracket 0, N-2-i \rrbracket} \frac{\theta_0 - \theta_{N-1-j}}{1 - \theta_{N-1-j}} \quad (28)$$

If all the eigenvalues are assumed to be nonnegative, we get the identity

$$\mathcal{G}(\theta_0) = \sum_{i \in \llbracket 0, N-1 \rrbracket} a_i \mathcal{G}(\theta_i, \theta_{i+1}, \dots, \theta_{N-1})$$

If there are some negative eigenvalues, we end up with a result even more surprising. Let  $\Theta_M$  be the set of eigenvalues  $(\theta_i)_{i \in \llbracket 0, N-1 \rrbracket}$  in nonincreasing order of irreducible and reversible  $N \times N$  Markovian matrices. If  $(\theta_i)_{i \in \llbracket 0, N-1 \rrbracket}$  is an a priori given family of numbers with  $\theta_0 \in (0, 1)$  and such that  $(\theta_i/\theta_0)_{i \in \llbracket 0, N-1 \rrbracket}$  belongs to  $\Theta_M$  and if in (19) we put the weight  $a$  defined by (28), we obtain  $\mathcal{G}(\theta_0)[n] = (1 - \theta_0)\theta_0^{n-1}$ . Indeed, let  $\tilde{P}$  be an irreducible and reversible  $S \times S$  Markovian matrix admitting  $(\theta_i/\theta_0)_{i \in \llbracket 0, N-1 \rrbracket}$  as eigenvalues. It is then sufficient to apply (19) with the absorbing transition matrix  $\bar{P}$  naturally constructed from its  $S \times S$ -restriction  $P = \theta_0\tilde{P}$ .

Again the continuous time equivalent is simpler. Let  $\bar{L}$  be an absorbing generator satisfying our usual hypotheses and assume that there exists a constant  $a > 0$  such that

$$\forall x \in S, \quad \bar{L}(x, \Delta) = a$$

Necessarily  $a = \lambda_0$  and independently from the initial distribution  $m_0 \in \mathcal{P}(S)$ , the absorbtion time  $T_X$  is distributed as an exponential distribution  $\mathcal{E}(\lambda_0)$ . Computing the death proportions via (14), we obtain the identity

$$\mathcal{E}(\lambda_0) = \sum_{0 \leq i \leq N-1} a_i \mathcal{E}(\lambda_i, \lambda_{i+1}, \dots, \lambda_{N-1})$$

with

$$\forall i \in \llbracket 0, N-2 \rrbracket, \quad a_i = \frac{\lambda_0}{\lambda_i} \prod_{j \in \llbracket i+1, N-1 \rrbracket} \frac{\lambda_j - \lambda_0}{\lambda_j}$$

• To finish, we investigate the case  $N = 2$  in continuous time, to determine the set of laws of absorption time in this very simple situation.

So we consider the state space  $\bar{S} = \{0, 1, \triangle\}$  endowed with the generator matrix

$$\bar{L} := \begin{pmatrix} -(a+b) & a & b \\ c & -(c+d) & d \\ 0 & 0 & 0 \end{pmatrix}$$

with  $a > 0$ ,  $c > 0$  and  $b \vee d > 0$ , to insure the validity of assumption (B1). Note that (B2) is automatically satisfied here.

The eigenvalues of  $L = \begin{pmatrix} -(a+b) & a \\ c & -(c+d) \end{pmatrix}$  are

$$\lambda_0 = \frac{1}{2}(a+b+c+d - \sqrt{(a+b+c+d)^2 - 4(bd+cb+da)}) \quad (29)$$

$$\lambda_1 = \frac{1}{2}(a+b+c+d + \sqrt{(a+b+c+d)^2 - 4(bd+cb+da)}) \quad (30)$$

Let also be given a probability  $m_0 := (e, f)$  on  $S$ . Using the formula

$$m_0 \bar{L} = -\lambda_1 m_0 + \lambda_1((1-d_0)\mu_1 + d_0\delta_\triangle)$$

where  $\mu_1$  is a probability on  $S$ , it follows that  $m_0 \bar{L}(\triangle) = eb + fd = \lambda_1 d_0$ , i.e.

$$d_0 = \frac{eb + fd}{\lambda_1} \quad (31)$$

With these quantities, the law of the absorption time of the corresponding Markov process is given by

$$(1-d_0)\mathcal{E}(\lambda_0) * \mathcal{E}(\lambda_1) + d_0\mathcal{E}(\lambda_1)$$

Conversely, we wonder if for any real numbers  $0 < \lambda_0 < \lambda_1$  and  $0 \leq d_0 < 1$ , we can find  $a, b, c, d, e, f$  as above such that (29), (30) and (31) are satisfied. This is indeed true and even if we only consider birth and death processes, namely satisfying the restriction  $b = 0$ .

**Proposition 24** *Given  $0 < \lambda_0 < \lambda_1$  and  $0 \leq d_0 < 1$  there exist  $a > 0$ ,  $c > 0$ ,  $d > 0$  and  $f \in [0, 1]$  such that*

$$\begin{aligned} \lambda_0 &= \frac{1}{2}(a+c+d - \sqrt{(a+c+d)^2 - 4da}) \\ \lambda_1 &= \frac{1}{2}(a+c+d + \sqrt{(a+c+d)^2 - 4da}) \\ d_0 &= \frac{fd}{\lambda_1} \end{aligned}$$

### Proof

The two first equations of the proposition are equivalent to

$$\begin{cases} ad &= \lambda_0 \lambda_1 \\ a+c+d &= \lambda_0 + \lambda_1 \end{cases}$$



and this leads to the following second order equation in  $a$ :

$$a^2 + (c - \lambda_0 - \lambda_1)a + \lambda_0\lambda_1 = 0 \quad (32)$$

Its discriminant

$$(c - \lambda_0 - \lambda_1)^2 - 4\lambda_0\lambda_1 = (c - (\sqrt{\lambda_1} - \sqrt{\lambda_0})^2)(c - (\sqrt{\lambda_1} + \sqrt{\lambda_0})^2)$$

must be nonnegative, so we must have  $c \in \mathbb{R}_+^* \setminus ((\sqrt{\lambda_1} - \sqrt{\lambda_0})^2, (\sqrt{\lambda_1} + \sqrt{\lambda_0})^2)$ . Requiring furthermore that equation (32) must have a positive solution implies that  $c \in (0, (\sqrt{\lambda_1} - \sqrt{\lambda_0})^2]$  and for these values of  $c$ , we get two positive solutions (which is double for  $c = (\sqrt{\lambda_1} - \sqrt{\lambda_0})^2$ ),

$$\begin{aligned} a_-(c) &:= \frac{1}{2}(\lambda_0 + \lambda_1 - c - \sqrt{(c - \lambda_0 - \lambda_1)^2 - 4\lambda_0\lambda_1}) \\ a_+(c) &:= \frac{1}{2}(\lambda_0 + \lambda_1 - c + \sqrt{(c - \lambda_0 - \lambda_1)^2 - 4\lambda_0\lambda_1}) \end{aligned}$$

As  $c$  goes from 0 to  $(\sqrt{\lambda_1} - \sqrt{\lambda_0})^2$ ,  $a_-(c)$  (respectively  $a_+(c)$ ) goes from  $\lambda_0$  to  $\sqrt{\lambda_0\lambda_1}$  (resp. from  $\lambda_1$  to  $\sqrt{\lambda_0\lambda_1}$ ).

To end the proof, we note that the third equation of the above proposition permits to choose  $d_0$  in the range  $[0, d/\lambda_1] = [0, \lambda_0/a]$ , so all values of  $d_0$  in  $[0, 1)$  can be reached, since  $\lambda_0/a_-(c)$  converges to  $1_-$  as  $c$  goes to  $0_+$ . ■

The fact that birth and death processes lead to the same set of laws of absorption times as in the general case (always under assumptions (B1) and (B2)) is specific to  $N = 2$ , because for  $N \geq 3$ , we may have multiple eigenvalues in the general case, but not for birth and death processes. We wonder if for  $N \geq 3$ , the set of laws of absorption times of birth and death processes is dense in the set of general absorption times.

## Acknowledgments:

The starting point of this paper comes from a question of Persi Diaconis (one among several ten of equally motivating problems!), asking for a probabilistic interpretation of eigenvalues of Markov kernels. I'm furthermore indebted to Persi Diaconis for all his reading suggestions and support and to Jim Fill who has send his recent preprints [11, 12] on the subject. I'm also grateful to the ANR's Chaire d'Excellence program for providing financial support and to the Laboratoire J. A. Dieudonné (UMR 6621) in Nice, where this work was done, for its hospitality.

## References

- [1] David Aldous and Persi Diaconis. Strong uniform times and finite random walks. *Adv. in Appl. Math.*, 8(1):69–97, 1987.
- [2] David Aldous and Jim Fill. Reversible Markov chains and random walks on graphs. Monograph in preparation, available on the web site: <http://www.stat.berkeley.edu/~aldous/RWG/book.html>, 1994-2002.
- [3] Robert F. Botta, Carl M. Harris, and William G. Marchal. Characterizations of generalized hyperexponential distribution functions. *Comm. Statist. Stochastic Models*, 3(1):115–148, 1987.
- [4] Christian Commault and Stéphane Mocanu. Phase-type distributions and representations: some results and open problems for system theory. *Internat. J. Control*, 76(6):566–580, 2003.
- [5] Persi Diaconis and James Allen Fill. Strong stationary times via a new form of duality. *Ann. Probab.*, 18(4):1483–1522, 1990.
- [6] Persi Diaconis and Laurent Miclo. On times to quasi-stationarity for birth and death processes. Preprint, available on <http://hal.archives-ouvertes.fr/hal-00164690>, 2007.
- [7] Persi Diaconis and Laurent Saloff-Coste. Separation cut-offs for birth and death chains. *Ann. Appl. Probab.*, 16(4):2098–2122, 2006.
- [8] Jian Ding, Eyal Lubetzky, and Yuval Peres. Total-variation cutoff in birth-and-death chains. Preprint, available on <http://arxiv.org/abs/0801.2625v3>, 2008.
- [9] Patricia D. Egleston, Terry D. Lenker, and Sivaram K. Narayan. The nonnegative inverse eigenvalue problem. *Linear Algebra Appl.*, 379:475–490, 2004. Tenth Conference of the International Linear Algebra Society.
- [10] James Allen Fill. Time to stationarity for a continuous-time Markov chain. *Probab. Engrg. Inform. Sci.*, 5(1):61–76, 1991.
- [11] James Allen Fill. The passage time distribution for a birth-and-death chain: Strong stationary duality gives a first stochastic proof. Preprint, 2007.
- [12] James Allen Fill. On hitting times and fastest strong stationary times for skip-free processes. Preprint, 2007.
- [13] Qi-Ming He and Hanqin Zhang. Spectral polynomial algorithms for computing bi-diagonal representations for phase type distributions and matrix-exponential distributions. *Stoch. Models*, 22(2):289–317, 2006.
- [14] Qi-Ming He and Hanqin Zhang. *PH*-invariant polytopes and Coxian representations of phase type distributions. *Stoch. Models*, 22(3):383–409, 2006.
- [15] Samuel Karlin and James McGregor. Coincidence properties of birth and death processes. *Pacific J. Math.*, 9:1109–1140, 1959.
- [16] Tosio Kato. *Perturbation theory for linear operators*. Classics in Mathematics. Springer-Verlag, Berlin, 1995. Reprint of the 1980 edition.

- [17] Julian Keilson. Log-concavity and log-convexity in passage time densities of diffusion and birth-death processes. *J. Appl. Probability*, 8:391–398, 1971.
- [18] John T. Kent. Eigenvalue expansions for diffusion hitting times. *Z. Wahrsch. Verw. Gebiete*, 52(3):309–319, 1980.
- [19] John T. Kent. The spectral decomposition of a diffusion hitting time. *Ann. Probab.*, 10(1):207–219, 1982.
- [20] John T. Kent. The appearance of a multivariate exponential distribution in sojourn times for birth-death and diffusion processes. In *Probability, statistics and analysis*, volume 79 of *London Math. Soc. Lecture Note Ser.*, pages 161–179. Cambridge Univ. Press, Cambridge, 1983.
- [21] Charles A. Micchelli and R. A. Willoughby. On functions which preserve the class of Stieltjes matrices. *Linear Algebra Appl.*, 23:141–156, 1979.
- [22] Marcel F. Neuts. *Matrix-geometric solutions in stochastic models*, volume 2 of *Johns Hopkins Series in the Mathematical Sciences*. Johns Hopkins University Press, Baltimore, Md., 1981. An algorithmic approach.
- [23] Colm Art O’Cinneide. Characterization of phase-type distributions. *Comm. Statist. Stochastic Models*, 6(1):1–57, 1990.
- [24] Colm Art O’Cinneide. Phase-type distributions and invariant polytopes. *Adv. in Appl. Probab.*, 23(3):515–535, 1991.
- [25] Colm Art O’Cinneide. Phase-type distributions: open problems and a few properties. *Comm. Statist. Stochastic Models*, 15(4):731–757, 1999.

miclo@latp.univ-mrs.fr  
 Laboratoire d’Analyse, Topologie, Probabilités  
 Centre de Mathématiques et Informatique  
 Université de Provence  
 39, rue Frédéric Joliot-Curie